

Burning issues

The Kyoto Protocol is just a small first step in restricting man's influence on climate. If we can't prevent fires in Indonesia, such international efforts to limit the effects of climate change could be in vain.

The engagement of society's upper echelons with the problems of climate change reached a peak last week, when Queen Elizabeth II, accompanied in Berlin by ministers from Britain and Germany, launched a bilateral collaboration of the two countries' climatologists, business and investment communities.

While recognizing the newly energized Kyoto Protocol as a crucial first step in international collaboration, the meeting that followed focused on its limitations (see www.britischesbotschaft.de/statevisit/en/press/climate_change_conference.htm). As one participant said, the world cannot wait for a ponderous succession of interminably negotiated increments to treaties to be achieved: we have years, not decades, to take steps that might prevent the worst possible consequences of current trends in greenhouse-gas emissions later this century. As a UK House of Lords report emphasized this week, emissions trading needs to take aviation into account. Two other essential goals include protecting carbon sinks and integrating southern countries into the framework.

For a stark example of the latter challenges, consider the peatlands of Indonesia. Under the Kyoto Protocol, it is technically possible for investors to back projects that increase the capacity of ecosystems to absorb carbon dioxide from the atmosphere, and to sell the resulting 'carbon credits' to polluters who need more time to control their emissions — though no such projects have yet been approved by the United Nations. The idea, which will be explored in detail in a News Feature in *Nature* next week, is to harness market forces to help limit atmospheric levels of CO₂. But success is far from guaranteed, and ecologists who study the world's peatlands are already pointing to a major gap in the Kyoto Protocol's carbon-trading provisions.

As soon as peat is drained, microbial activity previously held in

check by waterlogging causes it to start releasing CO₂. If large expanses of drained swamp catch fire, the gas belches into the atmosphere in staggering quantities. Yet until 2012 at the earliest, no one can claim carbon credits under Kyoto for restoring the hydrology of a damaged peatland ecosystem — unless the project can somehow be shoe-horned under the headings of forestry or agriculture.

Anyone who doubts the seriousness of this oversight should consider what happened in 1997, when El Niño conditions brought months of drought to southeast Asia. Millions of hectares of drained Indonesian peat swamps went up in smoke, releasing perhaps as much CO₂ as Europe emits each year by burning fossil fuels. The next time a severe El Niño hits, this catastrophe will be repeated (see page 145).

Negotiators considering the shape of the Kyoto Protocol after 2012 clearly should include peatland restoration under its carbon-trading arrangements. But we can act before then. The onus lies with Indonesia itself and the nine other members of ASEAN, the Association of Southeast Asian Nations. They collectively lost billions of dollars in the choking haze that smothered the region in 1997, and have set up a task force to consider the issue. But ASEAN has been slow to turn its concern into practical projects designed to prevent fires in Indonesia's peatlands.

As for the rest of the international community, little action has been taken, save for some tiny pilot projects to block the drainage canals that created the tinderbox in the first place. There is a real opportunity here to spend development aid money in a way that will legitimately benefit both recipient and donor. This would be a highly cost-effective way to simultaneously fix an ecological disaster and limit global warming. ■

The stem-cell state

California's citizens have changed the landscape of a key area of biology — with intriguing implications for everyone else.

The dream has come true for biologists in California who want to work with human embryonic stem cells. A large sum of money (\$300 million annually for ten years), the promise of new buildings, a state research institute dedicated to the field and a constitutional guarantee of the right to do the work all sailed through in a referendum last week (see page 135).

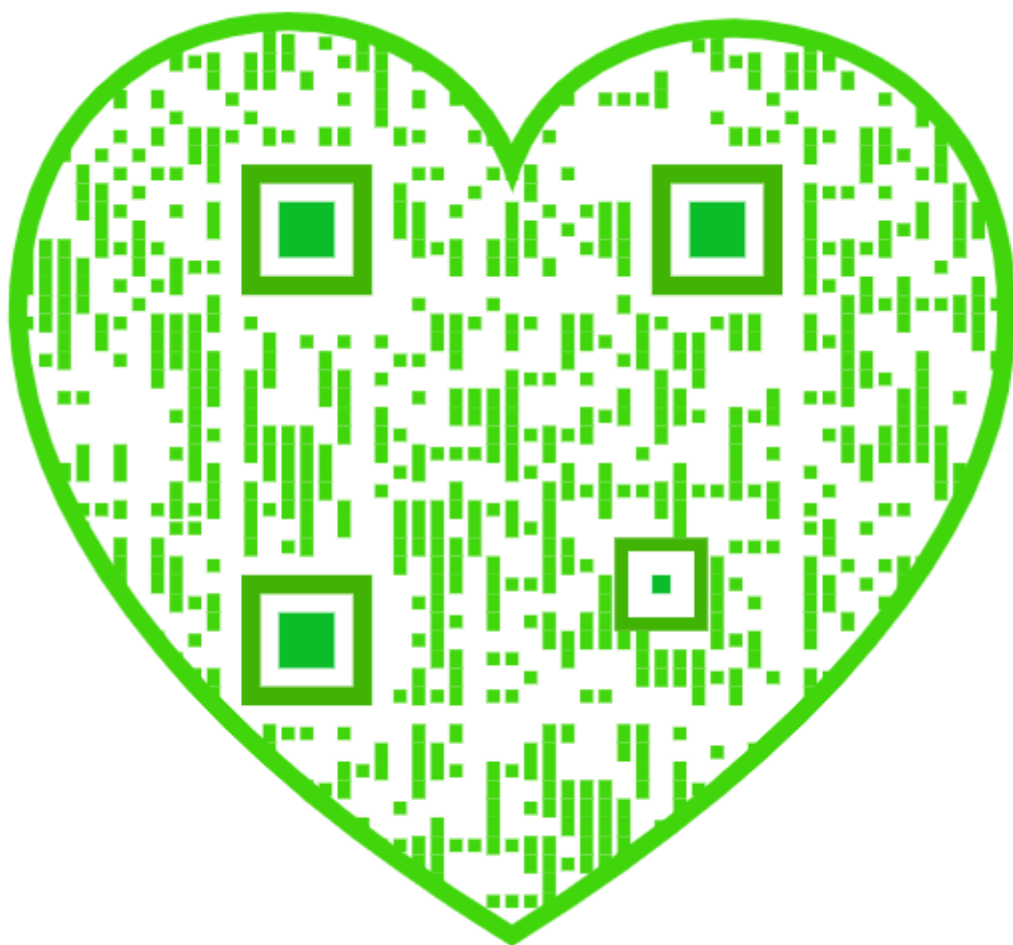
The passage of Proposition 71, as the measure is called, reflects the faith of the public in science's potential to make life better. The infrastructure it creates will make California one of the most suitable places in the world for pushing the frontiers of human embryonic stem-cell research. We have signalled a need for caution (see *Nature* 431, 723; 2004), because Proposition 71 is an unusual experiment in science funding. Nevertheless, many leading researchers have staked their reputations on its success and deserve credit for their hard work to pass the measure.

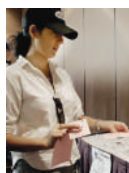
There is still a chance that President George Bush and his allies in Congress who oppose the research could undermine Proposition 71 with a federal ban on essential techniques such as 'therapeutic cloning'.

But such a bill has failed twice in the Senate already, and the chances that it would pass are less now than some prominent conservatives have lent their support to embryonic stem-cell research.

Researchers outside California will be right to worry that Proposition 71 could weaken embryonic stem-cell research elsewhere. Strong privately funded US research centres exist outside California — at the University of Wisconsin and Harvard, for instance — but young researchers especially are likely to feel the pull westwards, and even senior people may find it hard to pass up the lure of new cash and more lab space.

Current limitations on federal funding for the research mean that there are relatively few groups already working with human embryonic stem-cell lines in California. The administrators of the institute, who are to be appointed within 40 days, and the grant review board they will nominate, must avoid merely enhancing existing programmes and recognize the potential of newcomers to stimulate innovation. At the same time, Californian stem-cell research should strive where possible to grow from within. ■





Cell, cell, cell
California votes in
favour of stem-cell
research fund
p135



Safety test
Mouse experiments
give RNA interference
a shot in the arm
p136



Keeping track
Health agency plans
coordinated approach
to clinical trials
p137



Taste challenge
Chef seeks scientists
to cook up a storm
in the kitchen
p139

Watchdogs call academies to account over conflicts of interest

Tony Reichhardt, Washington

The National Academy of Sciences has changed the composition of an expert panel following charges of conflict of interest from environmental watchdogs.

The change in personnel for a group studying the disposal of toxic waste from the coal industry is indicative of the growing strength and sophistication of environmental and health advocacy groups, academy officials say. Several committees have been challenged in the past year alone, leading to resignations and shifts in panel membership.

The academy and its sister organizations, the National Academy of Engineering and the Institute of Medicine, perform about 250 studies a year for the federal government, assembling panels of experts to provide neutral advice on scientific and technical issues.

The Committee on Mine Placement of Coal Combustion Wastes met for the first time on 27 October to consider the question of how coal ash and other waste products should be disposed of in mines. Environmental groups say that the waste, which includes potentially toxic compounds, poses a health risk when it enters groundwater. At stake is whether its disposal should be regulated under the strict guidelines used by the Environmental Protection Agency (EPA) to monitor toxic dumps, or under existing regulations governing mining — as the coal industry would much prefer.

In a letter to the academy on 26 October, the Washington-based Center for Science in the Public Interest (CSPI) and 41 other groups charged that several of the 14 proposed members of the panel have significant ties to the coal industry. The academy has since asked Edward Green, a lawyer at the Washington firm Crowell and Moring who has lobbied for the mining industry, to step down.

Patricia Buffler, an epidemiologist at the University of California, Berkeley, has also resigned — although for scheduling reasons, according to the academy. Buffler has been an adviser to the Electric Power Research Institute in Palo Alto, California, which represents US utility companies.

At least one other member may yet be



Fouled up: this lake in eastern Kentucky was polluted by waste leaking from a Virginia coal mine.

removed from the committee, says William Colglazier, the academy's executive officer. He adds that all panels are provisional until questions of conflict of interest are settled early in the study process.

Following criticism in 1997, the academy has been reforming the process it uses to provide objective advice to the government (see *Nature* 390, 104; 1997). As one of several reforms, it began posting the names of proposed panel members 20 days before the start of a study, giving the public time to comment on them.

Keeping an eye

The CSPI has made a particular point of scrutinizing these lists. The centre's Integrity in Science project was begun four years ago, with funding from the Beldon Fund in New York, "to raise awareness about the role that corporate funding and other corporate interests play in scientific research", according to the group's website.

In January, the CSPI and the Natural Resources Defense Council (NRDC) questioned the make-up of an academy panel investigating the clean-up of a mining region near Coeur d'Alene, Idaho. In March, the CSPI challenged a panel on the EPA's

regulation of air pollution partly on the grounds that its chosen chairman, William Happer, a physicist at Princeton University, New Jersey, was well known for his scepticism on global warming. Happer later decided to resign for reasons unrelated to the challenge.

In May, California's two Democratic senators, Barbara Boxer and Dianne Feinstein, wrote to the academy to back the NRDC's protests about potential conflicts of interest held by two members of a committee studying perchlorate in drinking water. Another NRDC challenge, to a panel on toxicity testing of environmental agents, led the academy to acknowledge a conflict of interest for two members, although they were kept on the committee because of their unique expertise.

Colglazier says that such public challenges are welcome, and that the academy process ensures the integrity and neutrality of its studies. In the case of the coal-waste committee, the watchdog groups provided information that the academy's study directors were unaware of when assembling the panel. "Conflict of interest really says nothing at all about the person's morals, it's just a statement of fact," says Colglazier, who nonetheless acknowledges the embarrassment it can cause to scientists removed from a panel. ■

R. SIMPSON/AP

Trust gives warm welcome to open access

Jim Giles, London

The Wellcome Trust, Europe's largest research charity, has become the latest grant-giving body to throw down the gauntlet to academic publishers in the debate over open-access literature.

All papers reporting the results of research funded by the trust will in future have to be placed in a central public archive within six months of publication, the organization said on 4 November. The move could bring the trust into conflict with publishers, who often hold exclusive rights on the use of such material. This in turn could restrict researchers' choices about which journals they publish in.

But advocates of open access suffered a setback on 8 November when the British government rejected proposals for reforms favouring open access. The proposals had been made in July by the House of Commons Science and Technology Committee (see *Nature* **430**, 390; 2004).

In particular, the government rejected the committee's call that it should instruct its research councils to provide money so that scientists could meet author charges in open-access journals. "The government does not think it should intervene to support one model or another," it said in a formal response to the committee report, adding that it was "also not convinced that the 'author-pays' model is inherently superior to the current model".

But the Wellcome Trust says that it may now establish a European version of PubMed



The Wellcome Trust's new London HQ: the charity has set out plans for open-access publishing.

Central, a US database of biomedical research. Wellcome officials are already talking about this possibility to the US National Library of Medicine, which runs PubMed Central, but have not set a date for creating such an archive. The trust is preparing to set aside 1–2% of its total annual spend of £400 million (US\$740 million) to cover the costs of the archive and of uploading papers. The

version uploaded would not necessarily be the publisher's final version.

Researchers funded by Wellcome could find that the new rules create some difficult choices. Some publishing houses, such as Elsevier, which publishes more than 1,800 journals including *Cell* and *The Lancet*, do not currently allow any version of a paper they have published to be placed on a public archive other than on websites restricted to the author's research institution.

"This will put publishers and researchers in a difficult position," acknowledges Robert Terry, a senior policy adviser at the trust's London headquarters. But Terry believes that journals will modify their policies to allow papers to go to central archives. He points out that the US National Institutes of Health (NIH) is considering putting similar requirements on the research that it funds (see *Nature* **431**, 115; 2004). "It would be quite a strange journal that didn't include research funded by the NIH and the Wellcome Trust," he adds.

A spokeswoman for Elsevier said that the company was watching the NIH and Wellcome developments with interest but would not comment on possible changes to its copyright rules. Annette Thomas, managing director of Nature Publishing Group (NPG), which publishes *Nature*, says that important questions about the archive, such as who would take responsibility for the accuracy of the submissions, need to be addressed before NPG can take a position on the plan. ■

Leukaemia sleuths accuse state of nuclear cover-up

Barbara Simm, Munich

A 12-year-long investigation into a possible nuclear link to childhood leukaemia in northern Germany took a bizarre turn last week. On 8 November, six members of an eight-strong expert commission resigned, accusing the local authorities of covering up a fire at an alleged secret nuclear lab.

The startling accusation got prime-time media attention throughout Germany. State officials immediately rejected it as a "weird conspiracy theory", and announced that they would bring legal charges against Otmar Wassermann, a retired toxicologist at the University of Kiel who chaired the investigation.

The state government of Schleswig-Holstein in 1992 asked the commission to investigate the cause of an unusual cluster of leukaemia cases in the vicinity of the Krümmel nuclear power plant near Hamburg. Between 1989 and 1991, six cases of childhood leukaemia in the area

put incidence about ten times higher than the German average. Today, the number of new cancer cases has dropped, but is still above average.

In 1996, a report from the Öko Institute — an independent environmental research institute critical of nuclear energy — ruled out a link to a reactor operated by the nearby GKSS research centre (see *Nature* **384**, 398; 1996). Now the members of the Wassermann commission who resigned have added that the cluster cannot be blamed on the nuclear power plant either.

Wassermann claims to have found traces of radioactive products, including americium and plutonium isotopes, in soil near the GKSS centre. He says that the traces don't come from either Chernobyl fallout or from atomic bomb tests. Instead he suspects that the GKSS carried out "secret nuclear experiments" in the 1980s; a fire may then have released nuclear contamination that caused the cancer cluster, he says. An old

aerial photo of the area shows a building that Wassermann takes to be a secret lab.

Hans-Friedrich Christiansen, a spokesman for the GKSS, denies that any kind of secret or unauthorized experiments have ever been carried out at the centre, and says that there has never been a large fire. The building in the photo could have been a bunker from the Second World War, he adds.

Erich Wichmann, an epidemiologist at the National Research Center for Environment and Health in Neuherberg, and a member of Wassermann's commission who did not resign, says that Wassermann's conclusions are too speculative. Moreover, he says, independent labs have since failed to find unusual radioactive substances in the area.

A government spokesman says that Schleswig-Holstein has spent €4.5 million (US\$5.8 million) in 12 years to find the cause of the leukaemia cluster. The Wassermann commission was set to be wound up next February. ■

Joys match fears as California agrees to stem-cell proposal

Jonathan Knight, San Francisco

A \$3-billion programme that intends to make California a world centre for human embryonic stem-cell research got the green light from voters on 2 November, but some fear repercussions.

Proposition 71, which was approved by a healthy 18% margin, will raise the money by issuing bonds and will spend it on stem-cell research in the state over the next ten years. The win has caused jubilation among California biologists. But there are questions about what it will mean for US researchers outside the state — and about the possibility of retaliatory action by the re-elected president, George W. Bush, who opposes any research involving the destruction of human embryos.

The measure was conceived by several California film producers and businessmen with family histories of diabetes. They were frustrated by Bush's ban on federal funding for work on newly created embryonic stem-cell lines, which he implemented in August 2001. Their initiative gets around the ban by providing non-federal money and creating an Institute for Regenerative Medicine that will set priorities and draw up research guidelines.

"The model everyone has in mind is the National Institutes of Health," says Evan Snyder, who directs stem-cell research at the Burnham Institute in La Jolla, California.

But funding of basic research by a state on this scale is an experiment whose national implications could be profound. One concern is that Proposition 71 could weaken programmes in other parts of the country by luring talented researchers to California. Scientists will have good reason to relocate, says John Gearhart, a stem-cell researcher at Johns Hopkins University in Baltimore. "People will think: if there is money for ten years, why not California?" he says.

Beyond the money, there is also the allure of a single set of standards governing relevant procedures, such as egg collection. This will eliminate the need for researchers to negotiate research guidelines when collaborating with other California institutions, says Alta Charo, a bioethicist at the University of Wisconsin Law School, Madison.

But at the same time, Charo points out, it creates a disincentive for Californians to work with outsiders. "It's almost as if they have their own scientific country," she says. "California has a big enough system that there will be little reason to go outside."

In response to such pressures, other states may consider stem-cell measures of their own, suggests Kevin Wilson, a spokesman for the American Society for Cell Biology, based in Bethesda, Maryland. "There could be a domino effect," he predicts.

In Washington state, for instance, legislation that affirms the legality of working with human embryonic stem cells is gathering support. "There is certainly more interest in passing it," says Steven Gilbert, director of the Institute of Neurotoxicology and Neurological Disorders, a Seattle-based, non-profit group that promotes neuroscience research.

President Bush's opposition to the creation of new stem-cell lines from human embryos became a central issue in his successful re-election campaign, and the California initiative is in large part a reaction to his policies. It is not clear how, if at all, the president and his supporters will react to the measure: Republicans might try to get Congress to ban 'therapeutic cloning' outright.

Wilson says it is too soon to know whether such a move will occur. He says his society still hopes that the Bush administration will loosen restrictions on funding.

The first steps to implementing Proposition 71 were taken this week, but observers say it may be months before requests for research proposals are issued. A supervisory committee is due to be appointed by 13 December, and various elected officials, including the state's governor, will select most of the panel's 29 members from research institutions, private companies and disease advocacy groups in the state.

Research institutions around the state are already planning projects that might merit funding, says Snyder.

Bush set to keep core science team for second term

Emma Marris, Washington

George W. Bush's re-election on 2 November and his party's increased clout in Congress leave him in an even stronger position to set the national agenda on research, say science lobbyists.

The administration's scientific A-team looks set to stay — Elias Zerhouni is likely to remain the director of the National Institutes of Health and Bush appointee Arden Bement is on course to run the National Science Foundation (NSF).

The day after the election, Paul Gilman, head of research and development at the Environmental Protection Agency, said that he was leaving for a position in the private sector. The agency's chief, Mike Leavitt, has only held his post for a year, but some agency officials suggest that he may now move to a cabinet-level position, possibly as secretary of the interior.

Bush's energy secretary, Spencer Abraham, is also likely to move. "I would be very, very surprised if he stuck around for another four years," says one physics lobbyist. An *Associated Press* report suggests that Abraham may take Norman Mineta's job as transportation secretary.

Bush's science adviser, Jack Marburger, has given no indication of his plans.

Bement, the director of the National Institutes of Standards and Technology, was selected by Bush in September to lead the NSF. Formal nomination and confirmation by the Senate should come quickly, allowing Bement to assume permanent control of the agency, which he has been running as acting director since February.

Zerhouni, who is two-and-a-half years into a six-year term, is widely expected to stay on. His boss, the health secretary Tommy Thompson, has said in interviews that he may leave, although no official announcement has been made.

And observers of NASA predict that its administrator, Sean O'Keefe, will stay — at least until the space shuttle is flying again. John Logsdon, head of the Space Policy Institute at George Washington University in Washington DC, suspects that, after that, O'Keefe "would like another position, more related to national security".

Additional reporting by Geoff Brumfiel.



President Bush.

C. DHARAPAK/AP



Cast off: voters may have established California's research autonomy.

F. PROUSER/REUTERS/NEWS.COM

Herbicide critic dropped from pollution conference

Rex Dalton, San Diego

A California biologist is accusing Minnesota officials of censorship for blocking his conference lecture about environmental damage associated with a popular agricultural herbicide.

Tyrone Hayes of the University of California, Berkeley, says the Minnesota Pollution Control Agency was succumbing to pressure from pro-industry government officials when, late last month, it withdrew an invitation for him to deliver the plenary talk at its Air, Water and Waste Conference next February.

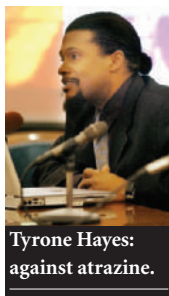
Hayes has published extensively on links between low levels of exposure to the herbicide atrazine and developmental defects in amphibians. He also cites studies linking atrazine to prostate cancer and decreased sperm counts in men. Syngenta, the Swiss-based firm that makes the herbicide, says there is no evidence of environmental damage.

Last year, the Environmental Protection Agency (EPA) re-registered atrazine despite a regulatory review that drew flak from Hayes for relying on too many industry-funded studies. He repeated his criticism on 25 October at a hearing of a committee of Minnesota state representatives, who are considering banning the herbicide.

Sheryl Corrigan, the pollution agency's commissioner, said in a statement that she doubted whether Hayes's research "was the right focus" for the conference. She instructed her staff to withdraw their invitation.

Hayes says that agency officials first asked him to cut the word 'atrazine' from his talk, which he refused to do. He adds that they told him that an EPA official "instigated" the "uninvitation" by contacting officials at the Minnesota Department of Agriculture, who then contacted the pollution agency. Agriculture officials acknowledged contacting the agency. EPA officials did not respond to interview requests.

After criticism from environmental groups, the pollution agency reinvited Hayes to give some form of lecture — not a keynote address. But Hayes says he doubts he will attend, because Corrigan blocked her staff from listening to his testimony. Corrigan claims that limiting staff attendance at hearings is routine. ■



Tyrone Hayes: against atrazine.

Hopes rise for RNA therapy as mouse study hits target

Erika Check, Washington

The efficacy of a promising therapeutic technique has been demonstrated in a study that might also relieve concerns about the approach's safety, researchers say.

Biologists at the company Alnylam Pharmaceuticals this week publish results showing that RNA interference — RNAi, for short — can lower cholesterol levels in mice (see page 173). The study is particularly encouraging because the treatment was injected directly into the bloodstream — a delivery method whose simplicity appeals to clinicians.

Human trials of RNAi, which aims to switch off disease-causing genes, have already begun for the eye disease macular degeneration. But these apply treatment directly to the eye. The Alnylam study, by introducing treatment through the bloodstream, could greatly increase the clinical value of RNAi, advocates say.

But RNAi specialists are also excited by aspects of the study that suggest the treatment did not interfere with genes other than those specifically targeted by the researchers. Experts in the field have been worried that the technique's clinical promise could run aground on unwanted 'off-target' effects. (A. L. Jackson and P. S. Linsley *Trends Genet.* **20**, 521–524; 2004).



Prick of the bunch: trials of RNAi therapy in mice could lead to treatments for humans with high cholesterol levels.

In RNAi, a cell is made to destroy pieces of RNA instead of translating them into a protein. The cell can be triggered to destroy pieces of this genetic material that match a particular target sequence, rather than simply shutting off all RNAs. As a result, RNAi is a powerful tool for silencing specific genes.

But work published last year found that RNAi could also change expression of RNAs that closely resembled a target RNA — not just the target RNA itself (A. L. Jackson *et al. Nature Biotechnol.* **21**, 635–637; 2003). Investigators have also found that RNAi can change the levels of proteins that are not related to the target RNA (P. C. Scacheri *et al. Proc. Natl Acad. Sci. USA* **101**, 1892–1897; 2004).

The Alnylam researchers, most of whom are based at the company's laboratories in Kulmbach, Germany, investigated these effects by measuring the levels of RNA expressed by genes that were unrelated to their target gene. They found that the expression of the unrelated genes did not change during the experiment. They also proved that the therapeutic effect — an overall decrease of cholesterol levels in the blood — was due to an on-target effect, by showing that the target RNA was destroyed at the points they hoped to hit with their treatment molecule.

"We included all the appropriate controls explicitly because we wanted to address the concerns people would have about off-target effects," says John Maraganore, Alnylam's chief executive, based in Cambridge, Massachusetts. He adds that the group's demonstration of how the mechanism actually cut cholesterol levels is the most definitive proof yet that RNAi works the way it is predicted to in mammals.

But Aimee Jackson of Rosetta Inpharmatics, a bioinformatics company based in Seattle, Washington, points out that further questions need to be answered before the technique moves to the clinic. The Alnylam team only analysed a few off-target genes, and did not analyse levels of many proteins from different organs, so its results do not prove that no off-target effects occurred. The mice were also studied over a relatively short period of time. "We don't know if there are effects on other organs," says Jackson. ■

Creative Commons ponders share options

Paula Gould, London

A not-for-profit organization is preparing to launch a form of science licensing that it says will give researchers more flexibility when they publish and share data.

The project, called Science Commons, has grown out of the Creative Commons movement, a scheme devised by Lawrence Lessig of Stanford Law School, California, to promote the online publishing of audio, visual and textual materials with "some rights reserved".

Science Commons aims to provide a form of legal protection that could serve as an alternative to both copyright and patents. If successful, the system should allow the creators of a pesticide, for example, to restrict its free use to the developing world through one simple licence, rather than a web of international patents. Most would declare this a worthy goal, but sceptics say it will be a hard slog for Science Commons, as those involved have little experience of patent law.

Creative Commons licences are free to use and legally binding. To date they have garnered most support from musicians and web loggers who wish to promote their work over the Internet, but who do not want to lose all control over its use. The movement's activities are funded primarily by three US-based private foundations, and are



Rights issue: Science Commons would provide legal protection for work published on the web.

run from premises at Stanford Law School.

Since its inception, the movement's founders have wanted to expand into the world of science. Additional funding to do so has now been obtained from an unnamed source. John Wilbanks, a fellow at the World Wide Web Consortium, an organization that aims to promote the development of the web, has been appointed director of Science Commons. He plans to consult with scientists, companies and funding agencies to work out a mechanism by which the commons will work. "We are not coming in with a pre-written agenda. We only want to solve areas of legal friction that the scientific community tells us are a problem," Wilbanks says.

The "some rights reserved" philosophy has already made inroads into the world of

science. A Creative Commons licence covers the content of the Public Library of Science publications *PLoS Biology* and *PLoS Medicine*. And the Biological Innovation for Open Science (BIOS) initiative, run by Richard Jefferson, aims to make methods and techniques developed by scientists freely available, in return for the results gained through such techniques also being freely released (see *Nature* 431, 494; 2004). Science Commons says it hopes to cover all this ground.

Wilbanks is in discussions with BIOS to explore possible link-ups. But Jefferson is sceptical of the impact that Science Commons will have outside the publishing arena. "The world of patents and science has almost nothing to do with the world of copyright. The economics, the culture and the pragmatics have almost no parallels," he says.

Science Commons will initially focus on biomedical sciences when it launches in January 2005. But Wilbanks would ultimately like to see the concept used in a wide range of scientific fields, including astrophysics and high-energy physics, where large amounts of data are collected. "The goal is to be an international Science Commons, not a US-centric life-sciences commons," he says. ■

► <http://creativecommons.org/projects/science/proposal>

WHO seeks system for tracking global clinical trials

Erika Check, Washington

The World Health Organization (WHO) hopes to earn international support next week for a far-reaching plan to set up a global tracking system for clinical trials.

The WHO will solicit support for the tracking system at its Ministerial Summit on Health Research, to be held in Mexico City on 16–20 November. WHO officials are proposing to establish an Internet portal that will give easy access to clinical-trial registries around the world. They also want to create a unified system for assigning unique identifiers to trials.

These measures are intended to reduce the duplication of trials and make it easier for medical researchers, the public, journal editors and governments to track them. By taking the lead on a global registry, it also hopes to set minimum standards for what information should be included in clinical-trial registries.

"This is a good idea that's been languishing for years and has been revitalized in the past six months," says Timothy Evans, assistant director-general for evidence and information policy at the WHO in Geneva. "Given the groundswell

of interest, we're very keen to get this trials registration into global practice as soon as possible."

Patient advocates and health researchers have long pushed for clinical-trial registries, saying they provide a fuller picture of the performance of drugs and devices in trials than that given by the often-favourable results published by manufacturers.

Registries have only recently started to receive strong support, however, following scandals such as the year-long controversy over the possible increased suicide risk among children taking antidepressants called selective serotonin reuptake inhibitors (SSRIs), which raised the profile of the issue. The US Food and Drug Administration said last month that SSRIs increase the risk of suicide in children, but based its conclusions on unpublished data that had been unavailable to the public.

"The reason this has caught on now is the SSRI issue," says Kay Dickersin, an epidemiologist at Brown University's Center for Clinical Trials and Evidence-based Healthcare in Providence, Rhode Island, who has been working with the WHO to



Under observation: a global registry could prevent the duplication of clinical trials.

develop plans for a global registry. "It has captured the public imagination."

But implementation of the WHO's global tracking system would face many obstacles, including financing — particularly for the incorporation of developing countries' clinical-trial data. The organization will find out in Mexico City next week how many nations back the concept. ■

Foreign-student enrolment declines in United States

Washington Enrolment at US universities by students from other countries fell in 2003–04 for the first time in more than 30 years.

The nation's largest annual survey of foreign students, carried out by the Institute of International Education in New York, found that international enrolment dropped by 2.4% compared with 2002–03 to 573,000

students. The institute says this is the first decline since the 1971–72 academic year.

The results add to a survey earlier this autumn showing that the number of foreign science graduate students admitted to top US institutions was down by nearly 18% (see *Nature* **431**, 231; 2004). The new survey includes both university and postgraduate students in all fields.

Much of the drop is due to falls in enrolment from China, Japan and Taiwan, the institute says, but there was a 7% increase in Indian students to 79,000. Enrolment from largely Islamic countries such as Indonesia and Pakistan continues a trend of decline that began in 2001, the year of the 11 September terrorist attacks.

Australia plots course to rights on marine microbes

Sydney Genome researcher Craig Venter has signed a deal with the Australian government allowing him to trawl its seas for novel microbes, sequence their DNA and publish the resulting data. In return, Venter has promised to attach a 'tagline' to the data stating that Australia retains the right to benefit from any commercial projects based on the information.

Venter announced this spring a plan to travel the world for two years collecting

marine microbes. He has struck deals with some governments to regulate this collection. In Bermuda, his activities highlighted loopholes in the law that meant he did not need a government permit (see *Nature* 429, 600; 2004). The Australian deal is the first to use tagging as a way for a government to retain commercial rights, he says.

Venter plans to deposit the sequence data in the publicly accessible GenBank, provided it allows the commercial tagline to appear on the website. "We'll create a whole new public database if GenBank refuses," says Venter.

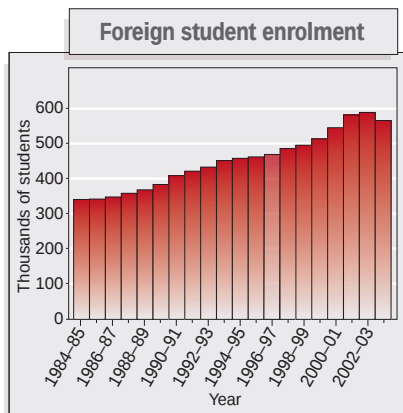
Physicists see fraud as colleagues forge ahead

Washington Young physicists are quite likely to have come across fraud in the course of their work, according to a survey carried out by the American Physical Society (APS).

The study, which appears in this month's issue of *Physics Today*, surveyed some 750 junior APS members. It found that 39% of them had seen their peers behave in an unethical way. The most common transgressions were including an undeserving author on a paper, or omitting a deserving one. But the group also reported incidents of plagiarism, data falsification and "less than truthful reports".

Although the self-selecting web-based

SOURCE: INST. INT. EDUCATION



Down-turn: the number of foreign students enrolling at US universities has taken a dip.

survey is unlikely to give an accurate picture of behaviour among all physicists, the results are still “startling”, says Kate Kirby, a physicist at the Harvard-Smithsonian Center for Astrophysics in Cambridge, Massachusetts, and part of an APS taskforce on ethics.

When asked what they thought motivated the dishonest actions, the respondents cited the pressure to publish in top-tier journals and the demands made on graduate students by senior faculty members.

Kirby says that the APS has responded by adopting new ethical statements on the treatment of subordinates. It is also considering other changes such as developing standards for keeping records of research.

Transgenic crops win food fight in California

San Diego Ballot propositions to ban the planting of genetically modified (GM) crops were defeated in two agricultural counties in California last week, after fiercely fought political campaigns.

Several Californian counties hosted a measure to ban GM crops in the 2 November presidential election, which led to months of heated debate. The anti-GM drive accused the pro-GM Farm Bureau of running a “dirty” campaign and of lying in its leaflets. The Farm Bureau, meanwhile, accused

Chef hungry for researcher with a taste for adventure

London Wanted: PhD student with a taste for more than pizza, pasta and beer. British chef Heston Blumenthal (right), best known for serving up snail porridge and bacon-and-egg ice cream, is seeking scientific assistance with his menu.

Blumenthal is hoping that a three-year investigation into novel flavour release mechanisms and hydrocolloids — gel-like substances that largely determine the texture of some foods — will yield new and even more remarkable treats.

The unusual studentship is being supported by the Biotechnology and Biological Sciences Research Council, which is providing a £13,000



(US\$24,000) annual stipend. Basic bench work is to be conducted at the University of Nottingham and at Blumenthal's development kitchen adjacent to his award-winning Fat Duck restaurant in Berkshire.

anti-GM speakers of giving “deceptive” talks on the downsides of the technology.

San Luis Obispo County, where vegetables and grapes are grown, and Butte County, a rice- and tree-growing area, both voted against the ban. A ban was approved only in Marin County, a region north of San Francisco with little agriculture. In Humboldt County, a similar measure failed when its wording was declared to be legally flawed.

The battle was not the first food fight to hit California. Mendocino County banned GM crops last March, despite a \$600,000 campaign by biotech and farm organizations in support of the technology.

Screening for malaria
Nature's supplement “Malaria: the long road to a healthy Africa” (430, 923–945; 2004) was produced in collaboration with the independent film production company Films of Record. Two related documentaries will be broadcast on BBC4 at 9 pm GMT on 17 and 18 November. “Malaria: fever road” highlights the experience of a Kenyan village fighting the illness with few resources. “Malaria: the vaccine challenge” covers the search for a vaccine and disparities in efforts between the developed and developing world. The films will also be broadcast in the United States on PBS next year.

A mole in hand...

A strange Australian mole has eluded scientific study for more than a century. Now biologists are teaming up with Aboriginal trackers to unearth the secrets of the itjaritjari. Carina Dennis checks on their progress.

A thin layer of red dust has settled across Joe Benshemesh's desk, coating boxes brimming with field instruments, wires, maps and notebooks. He has just returned from one of the most remote corners of Australia's outback. Still picking the dirt out from under his fingernails, he talks excitedly about the trip — his enthusiasm all the more remarkable given that he has seen the object of his studies only a few times.

Benshemesh, an ecologist at the Northern Territory Government's Biodiversity Conservation Division in Alice Springs, is studying the elusive sand-dwelling marsupial mole of Australia. The animal spends most of its life underground, rarely surfacing, and until now has been classified as 'too hard' to investigate by scientists. Even the local Aboriginal people — who have lived alongside the mole for thousands of years — know relatively little about its secretive life.

But they do know a lot about desert tracking. Now, in an unusual collaboration between scientists and an Aboriginal group, Benshemesh and his colleagues have combined ancient hunting methods with modern technology to learn more than ever before about this mysterious animal.

First recorded by Westerners in the late 1800s, the mole, often referred to by its Aboriginal name itjaritjari, caused a stir in zoological circles with its unusual features. At no more than 14 centimetres long, its tubular body fits into the palm of the hand. It lacks eyes and has only tiny holes for ears. And its short, cream-coloured fur, spade-like front claws and a backwards-facing pouch give it an unearthly appearance.

Despite the existence of a number of



Caught! An Aboriginal tracker proudly displays a rare find — the elusive itjaritjari.

pickled specimens in museums around the world, little is known about this creature. This is because the itjaritjari rarely surfaces, making observation in the wild virtually impossible. All specimens caught live have died within weeks in captivity for reasons that remain unclear.

Benshemesh's first job in finding new moles to study was to work out where to dig. For this he turned to the Aboriginal people of the Anangu Pitjantjatjara lands in central Australia. He sought from them both permission to track the mole, which lives mainly on their land, and help in finding the animal's hide-outs. Although the Aboriginals rarely track the mole for food — given that it's hard to catch and not much of a meal — they are familiar with the surface signs of all the local animals, including the mole.

Carved in sand

Benshemesh's key collaborator on the project is Robin Kankanakantja, an Aboriginal elder who lives at Walalkara, some 500 kilometres southwest of Alice Springs. Kankanakantja, who says he is well into his seventies, shares Benshemesh's enthusiasm about the mole, because it features in the

collection of Aboriginal spiritual beliefs known as *Tjukurpa* — sometimes translated as 'dreaming' — which still guides daily life for many Aboriginal people.

Although the two have become friends, the linguistic divide has been a challenge. "Our conversations can take a while — you need patience," says Benshemesh. The usual procedure for exchanging critical information about the mole begins with Benshemesh listening to Kankanakantja, who speaks a mixture of English and his native tongue, Yankunytjatjara. Benshemesh then repeats this back to him in English, and Kankanakantja corrects him as necessary.

In 1998, with Kankanakantja's guidance, Benshemesh began digging trenches in spots that showed evidence of mole inhabitation. Now, six years later, Benshemesh has learnt to recognize some signs of the mole, such as their tracks in the sand, but he has yet to develop a knack for consistently distinguishing the more subtle signs, such as small rises where tunnels have come near the surface. Yet the Aboriginal elders, after a lifetime of reading tracks in the bush, spot these almost intuitively.

One of the first discoveries to emerge from



Robin Kankanpakantja (left) has helped researchers to find where to dig for the marsupial mole.

J. BENSHEMESH

this joint search was that the itjaritjari produces stunning labyrinths of tunnels. This immediately dispelled one myth about the creatures: that their behaviour is like that of desert golden moles (*Chrysochloridae*) of southern Africa, which they strongly resemble. Desert golden moles are known as sand-swimmers because they produce no tunnels. The itjaritjari is clearly different. "If it was a sand-swimmer, you'd expect the sand to collapse behind it, leaving no trace of its path," says Benshemesh.

Blind vision

But more challenging questions that require modern gadgetry remain. Critically, how many of these creatures are there? To find out, Benshemesh set out 30 geophones, sensitive microphones that pick up vibrations in the sand, in a grid pattern over roughly a quarter of a hectare. He plans to eavesdrop on the moles' underground scratchings and use the data to estimate their speed and direction. With enough readings, it should be possible to estimate the size of the mole population in the region by comparing the rate at which moles tunnel with the extent of the tunnels determined by trenching.

Benshemesh would also like to learn more about the extent of the itjaritjari's distribution. To this end he is training Aboriginal people in the use of CyberTracker, a device originally developed in South Africa for ecological studies assisted by Kalahari bushmen. The hand-held unit allows the user to enter electronic notes while a satellite receiver monitors the user's location.

This approach alone could never completely scour Australia's vast network of sand ridges — stabilized dunes that cover some 1.2 million square kilometres. So Benshemesh has narrowed the search with the

help of geologist Bretan Clifford, an independent consultant who until recently worked at the Alice Springs base. Clifford has created a picture of the connections between individual sand ridges using digital maps derived from aerial photographs.

Assuming that an itjaritjari won't travel much farther than 250 metres on the surface because of its vulnerability to predators, Clifford has been able to break the ridges into six interconnected blocks that the mole is unlikely to move between. This allows information on the number and distribution of moles gleaned from one region to be extrapolated to the whole block. "It massively increases the efficiency of searching for the mole," says Clifford.

Finding more moles will be important to understanding its natural history. So far, only two species of itjaritjari have been described

— *Notoryctes typhlops* and *N. caurinus* — based on details of their appearance. DNA studies could shed more light on their diversity, but researchers have had to rely on preserved museum specimens, the DNA from which has usually degraded.

A recent breakthrough came when researchers discovered that they could get good DNA samples from the droppings of foxes, cats and dingos, all of which eat the moles (R. Paltridge *Australian Mammalogy* 20, 427–429; 1998). "The amount of marsupial mole that comes out of these dingo turds is huge and makes it really easy to study," says Steve Donnellan, a geneticist at the South Australian Museum in Adelaide.

From his studies of the droppings, Donnellan has preliminary evidence that there may be more than two species. "There is a hint of genetic variation there but it's not yet clear whether it's a different species or variation within species," he says. So far he has

"The amount of marsupial mole that comes out of dingo turds is huge and makes its DNA really easy to study."

— Steve Donnellan

looked only at differences in mitochondria, DNA-bearing compartments within the cells. He is now looking for a similar pattern of variation in genes in the cells' nuclei, which would suggest wider divergence and possibly speciation.

At the moment, if one is lucky enough to find an uneaten mole, recording its movements requires endurance. Only in the winter are daytime temperatures cool enough for fieldwork, but that means shivering through some very frosty desert nights. On one occasion, Benshemesh's team, many of them volunteers from the environmental group Earthwatch, managed to catch a mole on the surface. They released it and began recording from a geophone as it dived underground and started tunnelling. They took shifts over three days and nights to monitor its faint scratchings, producing one of the few existing recordings of itjaritjari.

Scratching around

Now that the geophones are in place, Benshemesh has plans to automate the recording and data analysis. For help, he turned to Don Simkins, a signal-processing expert with Zeta Associates, a company doing contract work for the US military base at Pine Gap in central Australia. The key is to figure out how to spot the sound of a mole in the din of background noise made by wind and other animals, and then to track its movement.

The more sophisticated the geophone monitoring becomes, the better it will be for eavesdropping on other mole activities, such as sex. "I'm really looking forward to that," says Benshemesh. He'd like to know not only how boy meets girl underground, but also how the young are reared. So far, he hasn't found any holes small enough to have been tunnelled by little ones. This means that the young either stay behind in a nest until they are grown or tag along behind their parents as they tunnel.

Benshemesh is impatient to monitor the animals underground because he suspects that those found on the surface are not behaving normally, and may even be sick. It's a risky place for a blind animal to be with foxes and dingos around, and they have no good reason to be there, he says. "My suspicion is that these are animals in trouble," says Benshemesh.

There is also a sense of urgency for the Aboriginal people who share the mole's land. Time is running out for the generation of trackers Benshemesh has worked with, and much of their knowledge is being lost. Western influence has left many of the younger generation feeling disconnected from their land and traditions. Benshemesh hopes that collaborative research on the mole will help to preserve indigenous culture. "It'll be a sad day when white fellas are teaching the Aboriginal people about tracking the mole," he says. ■

Carina Dennis is *Nature's* Australasian correspondent.

J. BENSHEMESH



Borneo is burning

Vast tracts of Indonesia's peat swamps have been drained in a misguided attempt to turn them into rice plantations. Now the landscape burns every year, belching smoke and hastening global warming. Peter Aldhous investigates.

Sitting in a small boat at the junction of two drainage canals, all I can see are walls of crumbling peat piled up to four metres high. On these parched banks, nothing grows. In some spots, smoke rises from the debris, while elsewhere patches of yellow, acidic sediments taint the peat. Overhead the Sun hangs orange-red, peering bleakly through a smoky shroud.

It's like a post-apocalyptic wasteland. Certainly 'apocalypse' is an apt description for what happened here in the Indonesian province of Central Kalimantan, in the south of the island of Borneo. This area was, until recently, a lush swamp forest that had lain undisturbed for thousands of years. But a single, wantonly inept decision by Indonesia's former dictator, Suharto, changed all that. Suharto wanted to turn Borneo into the rice bowl of Indonesia. But he succeeded only in creating a smouldering heap of ash that blights the lives of local people — and

threatens to destabilize the global climate by belching vast quantities of carbon dioxide into the atmosphere.

I'm here to meet the scientists and environmentalists who are assessing the damage, and are taking the first tentative steps towards a solution by blocking some of the drainage canals using hand-built dams. It will be an uphill struggle, but they hope to show that it is possible to stop the annual fires that choke and char this landscape, and to return the devastated swamps to life.

Much of the Indonesian archipelago is blanketed in a layer of peat — forest litter too wet to rot that has accumulated over thousands of years. Over decades, these forests have been slowly cleared and drained. But matters took a sharp turn for the worse in 1995. Fertile land on the overpopulated island of Java was needed for housing and industry, so Suharto announced that more than one million hectares in Central Kalimantan — an

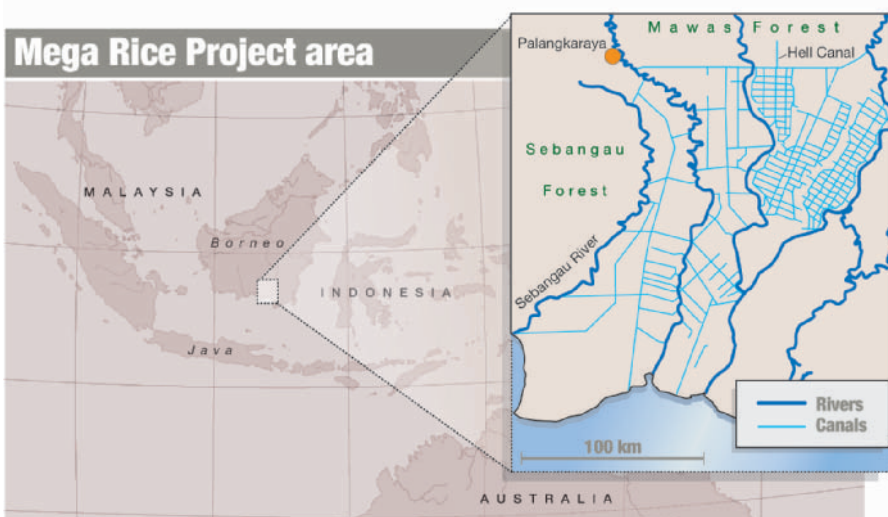
area about a third the size of Belgium — would be switched to rice production.

Over the next two years, loggers felled the forests, while contractors dug some 4,600 kilometres of drainage canals, the largest of them 30 metres wide (see Map, opposite). Tens of thousands of landless Javanese were brought in to tend the Mega Rice Project, as the plan was known.

Borneo's native Dayak people could only watch as their forests and the traditional livelihoods they had supported were destroyed. "First we were colonized by the Dutch, now we are colonized by Java," says Suwido Limin, whose team at the University of Palangkaraya, in the heart of the devastated area, studies the swamps. A fiercely proud Dayak, Limin was a vocal opponent of the plan from the start.

Sucked dry

The canals were laid down in a pattern that, in Java, keeps the soil well drained and irrigates crops with river water. But here in Borneo, the peatland topography rendered it useless. The peat is piled up into domed structures that rise to several metres above river level. Even a dictator's decree can't make water run uphill, so the canals simply sucked the peat dry. At the same time, the peat also proved too acidic to grow rice. Suharto had been told as much by the few local experts who were prepared to speak out. But the plan



Canals carved through Kalimantan (above) have wrecked its peat swamps. Now locals are trying to dam the waterways and restore the area (below).



went ahead anyway — buoyed by the lure of lucrative construction contracts doled out to associates of his regime.

In the end, the Mega Rice Project yielded barely a grain. Its failure occurred against a background of economic meltdown and rioting across much of Indonesia. In Kalimantan, there was an orgy of bloodletting as the downtrodden Dayaks turned on Muslim settlers. Some returned to the gruesome headhunting practices of earlier generations, invoking the spirits of their ancestors in pre-battle rituals. The newer migrants brought in for the Mega Rice Project mostly avoided the violence. But with no economic incentive to stay, many fled Borneo.

Today, Kalimantan's tensions have subsided. Suharto was forced to step down in 1998, and the Mega Rice Project was abandoned shortly afterwards. But the drainage of the swamps has left a legacy of fire that returns each year during the dry season, from July to late October. Some fires are started deliberately to clear land for cultivation; others result from carelessly tossed cigarettes.

This October, street life in Palangkaraya, the capital of Central Kalimantan, is accompanied by the constant whiff of smoke. Drive outside the city's limits, and you are soon in a smoggy haze. There are few flames, but the peat can smoulder for days. At one point, a fallen tree blocks the road — the peat in which it grew has literally burned away. Standing by the roadside, my eyes are streaming.

Burnt out

But this isn't a particularly bad year. Things really go to hell during the reversal of Pacific currents known as El Niño, which brings drought to the region. In 1997, the strongest El Niño on record encouraged fires that cut visibility in Palangkaraya to less than ten metres for almost three months. "Many people went to hospital with lung problems," recalls Alue Dohong, a local environmentalist. The city's airport was closed, and traffic on Kalimantan's rivers — the lifeblood of the area's struggling economy — was disrupted as boats collided in the smog.

Fires burned out of control across Indonesia for months. The haze extended across southeast Asia, and cost more than US\$4.5 billion in lost tourism and business¹. The burning peat resulted in the largest annual increase in levels of carbon dioxide in the atmosphere since records began in the 1950s (ref. 2).

According to satellite data analysed by Annette Bechteler, a student at the University of Munich in Germany, the 1997 disaster torched more than 2.7 million hectares in Central Kalimantan. And we can expect more of the same. In 2002, a weak El Niño saw the fires return. Given that the peat is more than 12 metres deep in places, it will burn again and again, each time drought returns.

Even what's left of the swamp forest isn't safe. The Mega Rice Project has lowered the water table in neighbouring areas, and some of the remaining migrants it attracted have now taken to illegal logging. This makes the

forest vulnerable to fire, and threatens the survival of Borneo's last orangutans (see 'The orang's last stand', overleaf).

Most of what we know about Kalimantan's peat swamp forest, and the consequences of its destruction, comes from projects based at Limin's Centre for International Co-operation in Management of Tropical Peatland, or CIMTROP. In 1993, Limin and Jack Rieley, a peatland ecologist at the University of Nottingham, UK, began surveying an area of peat swamp in the Sebangau river basin, south of Palangkaraya. Today, their field site forms part of a 50,000-hectare 'natural laboratory'. You reach it by riding in a motorized cart on a rickety railway laid on stilts rising from the peat.

Up in smoke

Rieley and his colleagues — including his wife Sue Page, an ecologist at the University of Leicester, and Florian Siegert, a remote-sensing expert at the University of Munich — revealed the huge impact of Indonesia's fires on global climate. By combining satellite data and field observations, they estimated that the 1997 fires released 13–40% as much CO₂ as a typical year's global emissions from burning fossil fuels².

The Sebangau swamp has been absorbing CO₂ and storing it in peat for the past 26,000 years³. Now, one of the planet's most important carbon sinks is poised to become a major source of the gas. Even without burning, the prospects aren't good. As soon as a swamp is drained, bacteria begin oxidizing the dried peat, releasing its stored carbon. "Fire hastens a process that is going on anyway," says Page.

The key to reversing the damage is to raise the water table, and to plant saplings in the remoistened peat. My boat trip was with a team monitoring an initial attempt to do just this. The Climate Change, Forests and Peatlands in Indonesia (CCFPI) project involves the environmental groups Wetlands International and Wildlife Habitat Canada. Given US\$3.2 million over five years by the Canadian government, CCFPI is working with local people and officials to prevent fires in both Borneo and Sumatra — where large areas of peat swamp have similarly been drained.

In the northernmost of the Mega Rice Project's canals, CCFPI is building dams to try to keep the peat saturated. It's a tall order, given that the dams must be constructed by hand, using local materials plus supplies brought in by boat. And at best, CCFPI's efforts can only demonstrate what needs to be done on a vast scale. "If this model works, we can get more funds from other sources," suggests Faizal Parish, who heads the Global Environmental Centre, based just outside Kuala Lumpur in Malaysia, which has orchestrated support for CCFPI.

Each dam's frame comprises two walls of

"The 1997 fires released up to 40% as much CO₂ as a typical year's global emissions from burning fossil fuels."

logs driven vertically into the sediment and bolted together. Logs are lowered into position by a pulley system, and then pile-driven by people jumping up and down on a cross-beam. The dam is then lined with impervious textile, before being packed with sandbags filled with clay.

CCFPI has so far built seven dams, with mixed success. At our base camp, the dam across the main canal draining from west to east is holding up well, maintaining a head of water more than a metre high. Several kilometres to the west, on the 'Hell Canal' cut northwards into the relatively pristine Mawas Forest, the upper of two dams is doing even better: the water level upstream is about two metres higher than below.

But the lower of the two Hell Canal blocks is doing little to impede the remaining flow. Its builders apparently failed to clear its base of tree stumps and loose debris before piling in the sandbags. Another dam on the main canal has been breached, bowed by the weight of water bearing down on it. Local people have also removed sandbags to create a slipway for their boats. "Perhaps we need better construction guidelines and supervision," says Parish.

Limin, Rieley and their CIMTROP collaborators also plan to block some of the Mega Rice Project's drainage canals using similar



Drained: the canals cut through southern Borneo have sapped the peat swamps of life.

techniques. Construction of their first two dams will start in the coming weeks, using Finnish funding brought in by Jyrki Jauhiainen, an ecologist at the University of Helsinki. Further dams will be built as part of a European Union project on restoring tropical peatlands, which has just won a grant of €1.5 million (US\$1.9 million) over three years.

Gas check

The CIMTROP projects will not only monitor rising water levels, but will also track the effect this has on CO₂ emissions from the peat. A Japanese team from Hokkaido University already has a mass of data from automated CO₂ sensors. Jauhiainen will take similar measurements before and after his dams are built, to see if emissions are, as predicted, reduced by the retention of water.

If any of these efforts are to work, the people of Central Kalimantan must be brought on board. "It's very important to have a local community involvement," says Dohong, who is CCFPI's Kalimantan coordinator. CCFPI is providing loans to impoverished peatland villagers to buy livestock; if they plant tree seedlings, and protect them from fire, the loan won't have to be repaid. And because the trees include species that can be tapped for gum, they will provide another source of income once they mature.

Political commitment will be crucial, too. Samsi Kulu, head of the planning department for the district covering CCFPI's dams, is convinced of the need to block more canals. But he is frustrated by contrary policies enacted by the central government. Even today, its contractors are dredging some of the Mega Rice Project canals in a misguided attempt to 'rehabilitate' the area using methods suitable only for other landscapes. "It's just making the situation worse," laments Kulu.

One of the biggest fears of those trying to restore the damaged peat swamps is that their efforts could sink in the mire of Indonesian politics — now facing a shake-up with the inauguration of the country's first directly elected president, Susilo Bambang Yudhoyono, on 20 October. He will soon be presented with a plan recommending that some parts of the Mega Rice Project area continue to be farmed — perhaps for oil palms — whereas others are earmarked for conservation. But the details remain unclear.

In the meantime, the pilot efforts of CCFPI and the CIMTROP researchers can do little to prevent hundreds of thousands of hectares going up in smoke the next time El Niño returns. "It will take 20 years to fix this problem," says Parish. ■

Peter Aldous is Nature's chief news & features editor.

1. Glover, D. & Jessup, T. (eds) *Indonesia's Fires & Haze: The Cost of Catastrophe* (International Development Research Centre/Institute of Southeast Asian Studies, Singapore, 1999).
2. Page, S. E. *et al. Nature* **420**, 61–65 (2002).
3. Page, S. E. *et al. J. Quaternary Sci.* **19**, 625–635 (2004).
4. Singleton, I. *et al.* (eds) *Orangutan Population & Habitat Viability Assessment: Final Report* (IUCN/SSC Conservation Breeding Specialist Group, Apple Valley, Minnesota, 2004.)

The orang's last stand

Central Kalimantan still has large expanses of peat swamp forest that haven't been drained and cleared for agriculture. But even here negative effects of the Mega Rice Project are being felt. Out-of-work migrants have raided these forests for wood, digging ditches to float away the most valuable trees. This could spell disaster for the endangered orangutan, for whom these forests are the most important remaining habitat.

Orangutans (pictured) used to range freely all over Borneo and Sumatra. But as the most accessible forests have been felled, populations have become concentrated in the swamps. Today, there are thought to be about 57,000 orangutans left in the wild⁴ — with the largest single population, of 6,900 individuals, in the Sebangau peat swamp forest south of Palangkaraya.

Simon Husson and Helen Morrough-Bernard — both graduate students at the University of Cambridge, UK — say that there were as many as 13,000 orangutans in Sebangau in the late 1990s. But the population crashed in 2000–01, mostly thanks to illegal logging. Orangutans spend

much of their lives up in the trees, and are highly sensitive to disturbance. "They don't like being in an area in which the trees are falling down," says Husson. His and Morrough-Bernard's unpublished work suggests that the animals were forced into smaller areas of pristine forest, where food shortages caused many to starve.

Loggers in Sebangau are currently being dissuaded by a patrol team operating from the University of Palangkaraya. But the orangutan population won't bounce back any time soon — an adult female typically has only one baby every seven years. And, ominously, loggers are still operating in the Mawas Forest to the northeast, in which the Borneo Orangutan Survival Foundation is trying to establish a reserve.



Linear models can't keep up with sport gender gap

Will women runners ever overtake men at the Olympics? Don't hold your breath.

Sir—Women sprinters may one day overtake men, according to A. J. Tatem and colleagues ("Momentous sprint at the 2156 Olympics?" *Nature* **431**, 525; 2004). As the holder of the world and European high-jump records for women over 50, I must say that this statistical prediction has been greeted with much laughter in athletic circles. The authors show linear regression lines and state that this model does not differ significantly from nonlinear approaches. I'm prepared to believe that. But much criticism is possible on both the data set and the logic behind the model.

The data set is very small: only one

sprint result per Olympiad, with a large variation in results. Taking the mean of the best 10 per year provides 40 times more data, leads to much less deviation and clearly shows nonlinearity (see www.antenna.nl/weia/Progressie.html).

A logical critique goes like this: an athlete can improve greatly by training three times instead of twice a week and can improve further by adding a fourth training session, and so on — but each additional session will give less improvement than the one before. It follows that the sport as a whole will show a similar nonlinear improvement. When statistics, nevertheless,

point to linear development, there must be something wrong. Most likely the 'linear' graph in fact consists of more nonlinear parts. For example, one part for the period when athletes were adding ever more training sessions, one part for when they reached a ceiling in adding sessions (around 1980), one part for when drug users were filtered out, and so on.

In which Olympiad will the form of the real nonlinear development become clear? I dare not guess.

Weia Reinboud

Simon Bolivarstraat 87, NL 3573 ZK Utrecht, the Netherlands

Sprint research runs into a credibility gap

Sir—A. J. Tatem and colleagues calculate that women may outsprint men by the middle of the twenty-second century (*Nature* **431**, 525; 2004). They omit to mention, however, that (according to their analysis) a far more interesting race should occur in about 2636, when times of less than zero seconds will be recorded.

In the intervening 600 years, the authors may wish to address the obvious challenges raised for both time-keeping and the teaching of basic statistics.

Kenneth Rice

MRC Biostatistics Unit, Institute of Public Health, Forvie Site, Robinson Way, Cambridge CB2 2SR, UK

Biology students find holes in gap study

Sir—We are students aged 16–18 in a Texas high school. Our biology teacher Vidya Rajan asked us to comment on the paper by A. J. Tatem and colleagues (*Nature* **431**, 525; 2004); we believe the projection on which it is based is riddled with flaws.

The idea of women running faster than men — although not novel (see B. J. Whipp and S. A. Ward *Nature* **355**, 25, 1992; and Correspondence *Nature* **356**, 21, 1992) — is interesting, but one cannot draw these conclusions based on generalization by extrapolation. Tatem *et al.* used a domain of 104 years to extrapolate to a domain of 252 years. It is not logical to say that the first 104 years will have data with exactly the same regression as the next 148 years. Using similar reasoning in 1992, Whipp and Ward suggested that women would run the marathon faster than

men by 1998. This has still not happened.

In Tatem and colleagues' study, men were measured for 32 more years than women. This ignores the possibility that women might be reaching a plateau: had women's times been unexpectedly high before 1934, one could trace a decreasing rate of change for post-1934 Olympians.

Improvements due to the increase in numbers of women running are likely to level off as the rate of increase in participation slows down (see www.olympics.org.uk/olympicmovement/olympicissueswoman.asp).

Finally, both men and women may reach a physiological limit beyond which they cannot progress.

With these factors taken into consideration, the predictions made from the extrapolation seem less than sound.

Advanced Placement Biology Class

A&M Consolidated High School, College Station, Texas 77840, USA

Mind the gap: women racers are falling behind

Sir—Trend extrapolation can be an inexact science, especially in sport. A. J. Tatem and colleagues (*Nature* **431**, 525; 2004) suggest, counterintuitively, that a future woman may run faster than her male counterpart over 100 metres. It is worth noting that the 'fastest human on the planet' is usually the world-record holder for the 200 m, not the 100 m. For example, Michael Johnson's running velocity for his current world record 200 m in 19.32 s was 10.35 m s^{-1} , whereas Tim Montgomery's for his 100 m in 9.78 s was 10.22 m s^{-1} .

However, sports physiologist Stephen Seiler has analysed Olympic and world championship running results and found

that the mean performance gender gap in the world records has actually increased from 10.4% in 1989 to 11.0% now (C. Holden *Science* **305**, 639–640; 2004). This held for seven of the eight events from 100 m upwards. The exception was Paula Radcliffe's marathon, which narrowed the gap from a relatively vulnerable 11.9% to 8.4%. Hence, in general, the gap has widened during the past 20 years.

Nevertheless, Tatem and colleagues make the very good point that only a minority of the world's population of women has been able to compete. Were China and India, with their vast populations, to come fully onstream in track and field sports, they could bring with them statistical outliers of both sexes who would demolish current records. But it is likely that there would still be a gender gap in the range of 7–10% in favour of the biologically advantaged men.

N. C. Craig Sharp

Sport Sciences, Brunel University West London, Uxbridge, Middlesex UB8 3PH, UK

A. J. Tatem and colleagues reply — We appreciate the interest generated by this light-hearted piece.

We were, of course, careful to caveat predictions with "if current trends continue". But if we were to follow all the advice we have received, we would be both correct and incorrect to fit a linear model. If we were incorrect, we should instead have fitted a two-part spline, a lowess curve, not a lowess curve, a rational function of polynomials, a quadratic model (predicting times regressing back to 1900 levels by 2100), a cubic model and an exponential curve. We should have both removed and added points, and were both correct and incorrect in our use of confidence intervals!

Will the gender gap continue shrinking? We look forward to finding out.

The science of life and death

A look at the development of biological weapons and the threat they carry.

Biological Weapons: From the Invention of State-Sponsored Programs to Contemporary Bioterrorism

by Jeanne Guillemin

Columbia University Press: 2004. 256 pp.

\$27.95, £18.50

Malcolm Dando

In 2005, parties to the Biological and Toxin Weapons Convention (BTWC) will consider the “content, promulgation and adoption of codes of conduct for scientists”. The United States, following a report by the National Academies, *Biotechnology Research in an Age of Terrorism*, has already established a National Science Advisory Board for Biosecurity, charged with increasing the controls on biological research to prevent its misuse. There is an urgent need for an informed debate among life scientists to ensure that any such controls effectively prevent misuse but interfere as little as possible with peaceful civil research and development. The problem for many life scientists is that they are not conversant with the issues surrounding biological weapons, biowarfare, bioterrorism and associated topics, despite almost a century of biology and medicine being used in offensive bioweapons programmes.

In *Biological Weapons*, Jeanne Guillemin aims to provide a historical context to our present concerns. Her book, she says, “is about the twentieth-century incorporation of biological weapons into the arsenals of industrial states and its implications for present times, when new technologies and persistent political animosities may allow even more ominous threats than in the past”.

Guillemin divides her history into three periods: a phase covering the initial scientific understanding of infectious agents until the BTWC came into force in 1975; a second phase after the BTWC prohibited offensive bioweapons programmes, but when some nations, notably the Soviet Union, persisted illegally; and the current phase of grave concerns about bioterrorism and massive investment in biodefence.

This is a difficult story to tell straightforwardly, as Guillemin acknowledges, because much of the archival material is still unavailable to scholars, and there has been “an unusual degree of misinformation and even disinformation”. The book’s chapters nevertheless cover the essential elements of the story. There is discussion of the United States’ weapons programme during and after the Second World War, and the US rejection of biological and toxin weapons in the run-

up to the BTWC. Guillemin also covers the gruesome Japanese bioweapons programme and the use of these weapons in China. And she discusses the Soviet programme in the later years of the Cold War, including the ‘yellow rain’ accusations of biological weapons being used in Asia, and the accidental release of anthrax at Sverdlovsk in the Soviet Union. She also considers the lesser, and more recent, South African and Iraqi programmes, and current concerns and responses to bioterrorism threats.

Guillemin argues, quite reasonably, that a variety of restraints — notably custom and law, technological problems, a lack of military interest, government and public opinion, and fears of retaliation — have luckily prevented the widespread use of biological weapons over the past century. She also puts forward a credible argument that a wide range of integrated policies, including stronger international controls to pre-empt proliferation, will be required to prevent the future development of biological weapons. Scientists will be interested to note, however, that their fellows have been among the factors promoting the development of biological warfare in the past. In discussing potential agents, for example, Guillemin refers to a study by the eminent US scientists Theodor Rosebury and Elvin Kabat which, in 1942, assessed the advantages and disadvantages of some 70 agents for use as weapons.

An extensive table compares this with a series of later assessments through to 2001. Again and again she stresses that scientists and others “who believed in the future of biological weapons saw their potential for fulfilling the goals of total war, that is, for the mass killing or debilitation of enemy civilians”.

Chapter 2, on Britain’s bioweapons programme, should be required reading for anyone who doubts this point. Particularly revealing are the views and actions of Frederick Banting, the discoverer of insulin, in support of the development of biological weapons early in the Second World War.

Before the BTWC, scientists at least had the excuse that they were doing nothing illegal in developing biological weapons, but that doesn’t apply to participants in the huge Soviet programme after 1975. Guillemin devotes a chapter to this, including the illegal Biopreparat expansion, but because the evidence available today is fragmentary, there is an uneasy sense that there is much more to be discovered.

The book is largely concerned with the history of biological weapons developments, but the final three chapters are of particular interest because they discuss the implications of that history for today. How should we best assess and respond to the problem of bioweapons proliferation and the threat of bioterrorism?

For those who believe that everything



At risk: despite decades of discussion about controlling biological weapons, the threat is still with us.

changed after the terrorist attacks of 11 September 2001, and that only then did we realize the possibility of terrorists using weapons of mass destruction, Chapter 8 will be enlightening reading. It details the escalation of high-level US concerns over this issue in the 1990s and the ever-increasing budgets allocated to deal with it. The al-Qaeda attacks and the subsequent anthrax postal attacks "intensified Clinton-era policies already in place" for national security and defence against bioweapons, says Guillemin. What did change, of course, was the scale of the budgets — including those for biodefence.

Given that the Bush administration has rejected the decade-long effort to strengthen the BTWC with a verification protocol, and that some US biodefence projects could be perceived by others as crossing the boundary between defence and offence, the vastly increased US budgets for technological solutions provoke some awkward questions. The book does not dodge them. Guillemin notes, for example, that Project Bioshield, "a biomedical equivalent of Reagan's Star Wars" defence programme, promised "universal protection from biological weapons" but was faced with the uncertainty of the threat, the technology, and the organization of national vaccinations or other campaigns. Such budgets necessarily draw more and more of the biological community into biosecurity, placing them under the restraints on openness and the free exchange of information that this work involves.

Not surprisingly, this sane and sensible book ends by arguing for a more balanced approach in which the United States joins again with its international allies to redevelop a multifaceted, integrated set of policies against the malign misuse of life sciences. ■
Malcolm Dando is in the Department of Peace Studies, University of Bradford, Richmond Road, Bradford BD7 1DP, UK. He is co-editor, with M. L. Wheelis and L. Rozsa, of the forthcoming book Deadly Cultures: Bioweapons from 1945 to the Present (Harvard University Press).

The view from Budapest

Our Lives: Encounters of a Scientist

by István Hargittai

Akadémiai Kiadó; 2004. 264 pp. €30

Henryk Eisenberg

When asked whether he believed in extraterrestrial beings, physicist Leo Szilard replied that they were already in our midst: they were called Hungarians, and he was one. The implication was that they had colonized our planet. Readers of István Hargittai's *Our Lives* will certainly be left with the



Leo Szilard led the Hungarian invasion of the scientific world.

impression that they have, or the scientific world, at least.

Hargittai is a professor in Budapest who works on symmetry in chemical reactions. He has written some 25 books, many of them devoted to conversations with prominent scientists. In his latest book, *Our Lives*, each of the 19 chapters is centred on a Nobel laureate he has met. Interspersed with the stories of their lives and work are his own reflections and reminiscences of an eventful life spent with a wide range of friends and acquaintances. Hargittai was a child during the tragic years of the Second World War in Hungary, and his family, being Jewish, suffered persecution and endured life in a ghetto and a labour camp. His father was conscripted and was killed on the Russian front. From these circumstances arises one of the themes that permeate the book: several of his chosen Nobel laureates also never knew their fathers, or lost them in early childhood. This, Hargittai believes, may have been a formative factor in their lives.

Another recurring theme of the book is the experiences of Hungarian Jews during the war and in its aftermath. Hargittai and his mother were saved from deportation to Auschwitz by a deal that rescued many Hungarian Jews at the eleventh hour. They were given shelter until Hungary's leader, Miklós Horthy, belatedly changed sides. Hargittai's brother and other members of the family were less fortunate, and their sufferings are movingly described. One of Hargittai's Jewish friends who was deported to Auschwitz and survived was the chemist László Kiss;

the story of his experiences at the hands of both the Nazis and their Hungarian epigones, the Arrow Cross, never previously related, is well worth the telling.

Hargittai's encounter with the chemist Gertrude Elion led him to describe another illustrious Jewish woman, the Hungarian mathematician Vera Sós, who was also persecuted, and her friend, Paul Erdős, who got away. Other Hungarian scientists whose early lives were marked by harrowing experiences were George Olah and two who settled in Sweden. One was Hargittai's friend and patron, the biochemist Lars (previously László) Ernster; the other was George Klein, the distinguished tumour biologist, who arrived fatherless in Sweden.

Ronald Hoffman, the focus of another of Hargittai's chapters, was born in Poland, not Hungary, but the story of his survival as a Jewish child in occupied Europe is one of the most remarkable.

He made his way to the United States with no money and after a

late start emerged as one of the outstanding chemists of our times.

Many remarkable characters flit through these pages. Hargittai finds the human side of another Hungarian Jew, the abrasive and much disparaged Edward Teller, and discusses the political influence of Arthur Koestler. Hargittai has interesting things to say on science in the Soviet Union, which he experienced at first hand as a graduate student in Moscow, where he became acquainted with several distinguished chemists. He recounts the remarkable episode of the denunciation on ideological grounds of Linus Pauling's theory of resonance. He also seeks to right the injustice done to his compatriot Árpád Furka, who received too little credit for his central contribution to the conception of combinatorial chemistry. In 1967, Hargittai married the chemist Magdolna Vámbídy, who became his co-author on several books.

Many more scientists appear in *Our Lives*. Hargittai's aim is to blend the achievements of modern science with his own life in this turbulent period, taking in his family, the special place of Jewish and Hungarian scientists and thinkers in twentieth-century history, anti-Semitism and the terror of Nazi persecution. This is not an easy task in such a short book, and the components are sometimes difficult to disentangle. But the stories he tells have a great deal to offer to anyone interested in the broader aspects of science in our time. ■

Henryk Eisenberg is in the Department of Structural Biology, Weizmann Institute of Science, Rehovot 76100, Israel.

The art of seeing science

Leonardo

by Martin Kemp

Oxford University Press: 2004. 304 pp.

£14.99, \$26

Stefano Grillo

The poet Paul Valéry once described the impression created by Leonardo da Vinci's notebooks. At first glance it seems that Leonardo wrote about "the most varied subjects depending on his mood and the contingencies of the day", by working "at the service, in turn, of each of his numerous Muses". In most of the pages of Leonardo's notes, illuminating remarks on engineering, the sciences and the visual arts lie scattered, with no apparent order, among drawings, sketches and simple information about his daily life. So the first question to ask about Leonardo's work is whether it should be treated as myriad unrelated fragments, or whether this multiplicity is articulated by a method, giving it an internal unity. What was the relationship, if any, between Leonardo's art and his science?

Early authors, including his sixteenth-century biographer Giorgio Vasari, implied that Leonardo's excessive interest in science distracted him from his work as an artist. They thought this was why he created so few paintings, most of which he notoriously

left unfinished. At the end of the nineteenth century, a different picture emerged. Valéry, like his contemporary Gabriel Séailles (a professor of philosophy from Paris), insisted that there was a unifying method in his thought, despite the outward appearance of the notebooks. Leonardo's painting grew directly out of his understanding of nature. His works of art look so real and alive precisely because they were based on an understanding of the laws that govern natural phenomena. To draw a deluge, for example, Leonardo used his many observations of water flow and his thoughts on fluid dynamics to create one on paper, on nature's own terms. Science, far from being detrimental to his art, was as relevant to it as it was to the design of his machines.

This conception of Leonardo is at the core of Martin Kemp's new book, which the author describes as dealing with why "the *Mona Lisa* and the flying machine were, for Leonardo, the same kind of thing". Kemp has an unmatched ability to write about science and art with equal understanding, as readers of his *Science in Culture* contributions to *Nature* will be aware. This allows him to illustrate his insights into the mind of Leonardo with a series of concrete and penetrating analyses of his scientific studies and works of art. Kemp's aim is to equip the reader, through the examples he gives, with a means of fruitfully approaching Leonardo's work by always keeping in mind the unity of his thought.

Kemp's discussion of Leonardo's science

brings into relief several recurrent themes, such as his ubiquitous use of the law of proportionality, his use of geometry to grasp physical reality through visualization, and his belief in the functionality of all natural forms. At the root of Leonardo's science was the conviction that all knowledge must spring from visual observation. For Leonardo, seeing involved the intellect just as much as the eye: it was a way of using sight to penetrate natural phenomena with the mind, as magnificently demonstrated in his drawings. Kemp suggests that it is this way of 'seeing', in the sense of understanding, that makes Leonardo so fascinating today, a view I share wholeheartedly. It seems to me that whereas Galileo, the founder of the experimental method, astonishes us because we find his thinking so similar to that of the modern, mathematically based sciences, Leonardo fascinates us precisely because he understood nature differently — not in terms of predictive models, but through sight.

The picture of Leonardo that emerges in broad outline from this book is of a man who made important discoveries but was working wholly within the framework of aristotelian and medieval science. Eventually, Leonardo's empirical observations brought to light internal contradictions in these inherited systems. He became dissatisfied with them but did not propose a new framework within which he could interpret his findings. I know very few writers who have described this internal movement in Leonardo's intellectual development as clearly as Kemp does; Cesare Luporini's masterly 1953 study *La mente di Leonardo* is a comparison that comes to mind.

Kemp's remarkable gift of writing with great clarity about a subject that is far from simple makes this wonderful and concise volume suitable for the general reader. The tone of the book is personal, with the first and last sections reading almost as a novel. Kemp's admiration for Leonardo's intellectual integrity is evident, as is his respect for the piousness with which Leonardo devoted himself to the study of the laws of nature (which he saw as the manifestation of God's design). Throughout the book, Kemp's argument moves almost imperceptibly from one area of science to another, then from science to art and, in the final pages, from life to art in one continuous, uninterrupted flow. This perfectly mirrors Leonardo's way of thinking, his constantly seeing analogies between the most different phenomena, his "never looking at anything without thinking of something else". Kemp's style of writing is a masterly illustration of Leonardo's principle that all form should fulfil a function. ■

Stefano Grillo is at the University of Perpignan and PROMES/CNRS, Rambla de la Thermodynamique, Tecnosud, 66100 Perpignan, France.



Film

The story of life

After the international success — both artistic and commercial — of their documentary film *Microcosmos*, about the world of insects, French directors Claude Nuridsany and Marie Pérennou now present the history of the Universe in *Genesis*, released in cinemas in France and Germany in October.

From the Big Bang to the first simple and ordered forms of life, and then more complex creatures emerging from the ocean, the

familiar themes of evolution are given a poetic re-working. An African shaman structures the narrative with colourful metaphors for the difficult science. He speaks of the repeating cycle of birth, love and death — themes illustrated in the film's extraordinary footage, which somehow manages to anthropomorphize its subjects by implying that they indulge in romance.

Federica Castellani

♦ www.genesis-lefilm.com

Making waves

How a stroll in the park led to the beginning of quantum electronics.

Charles H. Townes

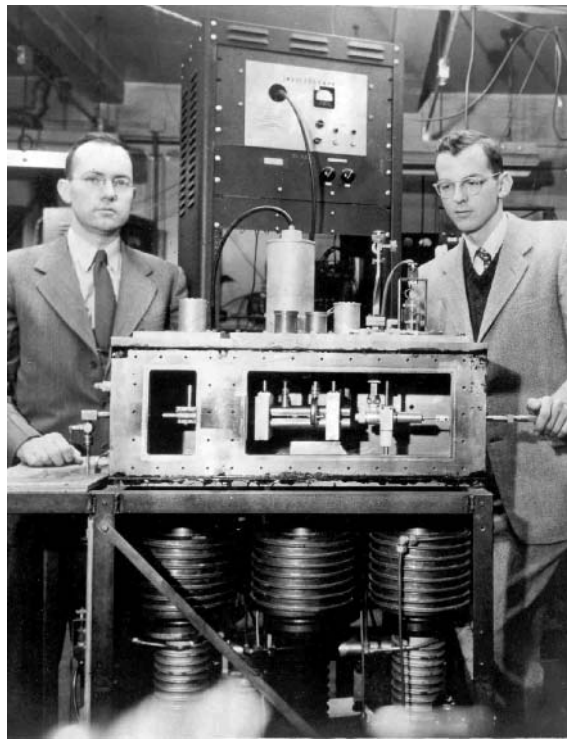
Immediately after the Second World War, physicists in several laboratories began working on the microwave spectroscopy of molecules. The field grew rapidly, but was soon given up by industrial labs — they saw no useful applications for the work — and moved to universities.

Almost everyone in the field wanted to obtain shorter waves, because the wealth of molecular lines and their intensity of absorption increased rapidly as one moved from centimetre to millimetre wavelengths, and then down into submillimetre or infrared wavelengths. Harmonic generation with electronic equipment could achieve millimetre wavelengths, but not the submillimetre range. At Columbia University, New York, I worked hard for several years looking for ways to obtain shorter waves, without much success. The US Navy was also interested in short waves, and in early 1950 asked me to form a national committee to search for ways of extending radar technology into the shorter wavelength region. We visited many centres and laboratories in the United States and Europe, looking for good ideas for producing short waves. None surfaced.

I called for one more meeting of our Navy committee on 26 April, 1951, in Washington DC. I was sharing a hotel room with Arthur Schawlow, a young postdoc from my lab, who was in town for an American Physical Society meeting. I woke up early in the morning, worrying about our lack of success in generating shorter waves. I didn't want to disturb Schawlow, so I left the hotel and went out to sit in nearby Franklin Park. It was a nice morning, but I was fretting over why we had found no solution for shorter wavelength oscillators. High frequencies needed a very small oscillator — hard to build and perhaps unable to take all the power needed to make it oscillate. Of course, molecules and atoms were small enough and easily produced very short waves. But to get a respectable amount of light out, high temperatures would be needed to excite sufficient numbers of molecules. Too high in fact — the molecules would all disassociate.

But wait a minute! This assumes that the molecules are in thermal equilibrium. What if they weren't? A collection of excited molecules in complete non-equilibrium would have no such limit to their potential radiation intensity. This was, at least, a possibility — but in practice would it produce an appreciable amount of power?

I pulled out a pen and an envelope and wrote down the necessary equations and



Charles Townes (left) and Jim Gordon with a beam-type maser.

numbers, using my favourite microwave molecule, ammonia, as an example. The most obvious way of obtaining many molecules in an excited state was to use molecular beam methods to separate those that were in the excited state from those that weren't. Of course, a resonant cavity tuned to the emission wavelength was also needed, to ensure that the emissions could be effectively captured and corralled. Great! The numbers said it could probably be made to work.

I raced back to my hotel room to tell Schawlow about the idea. He agreed, but didn't seem particularly excited. I decided to mull it over more before going public, so didn't mention it in the committee meeting.

Back at the lab, I decided to test the idea by first making an oscillator at longer centimetre wavelengths, where ammonia has intense resonances. If successful, I could then push into the submillimetre or infrared region. I began to look for a student who might give it a try and before long Jim Gordon, an outstanding student, turned up. With postdoc Herbert Zeiger we were set to start.

Two years later, Gordon and Zeiger had still not obtained oscillation. Polycarp Kusch, the departmental chairman at that time, and Isidor Isaac Rabi, his predecessor, came into my office and sat down. "Look Charlie," they said. "That experiment is

not going to work. We are in the molecular beam business and know it won't work. You know it won't work. You must stop, you are wasting the department's money." Rabi and Kusch were, of course, outstanding scientists — both received Nobel prizes. But as an associate professor, I had tenure. A department chairman could not fire me simply because of a disagreement or my incompetence. "No," I said. "I think the experiment will work and I am going to continue."

Two months later, Gordon dashed into the classroom where I was lecturing and excitedly declared: "It's working." The whole class left with me to go to the lab to witness the demonstration. We named this new kind of oscillator a MASER, for microwave amplification by stimulated emission of radiation. My students later suggested laser for

light emission, and iraser for infrared emission, although the latter was never used.

Most scientists at that time didn't believe the maser idea could be extended to such short wavelengths, because of the much higher decay rate of excited atoms or molecules as the wavelength became shorter. But I was sure we could go into the submillimetre or infrared range, and probably even to visible light. By the fall of 1957 I had figured out just how an optical maser (laser) could be built.

On a consulting visit to Bell Labs, I ran into Schawlow again and I told him about my ideas for an optical maser. I planned to optically excite atomic gas, have it radiate by stimulated emission, and use a cavity as a resonator. I wasn't completely happy with the cavity, as it would probably have multiple mode oscillations. Schawlow said, "Oh, I've also been wondering if that could be done," and suggested using two parallel mirrors, a Fabry–Perot, as a resonator. After a short delay while Bell Labs fixed up an appropriate patent, we published our ideas in 1958. There was immediate excitement. The maser had convinced industry that this was a valuable field, and very quickly there were many efforts to build a laser. The field of quantum electronics had begun. ■

Charles H. Townes is in the Department of Physics, University of California, Berkeley, California 94720, USA.

A cholesterol connection in RNAi

John J. Rossi

RNA interference — RNAi for short — might provide a way to silence disease-associated genes, but problems of delivery have hampered progress. Those problems may have been solved, at least in animal studies.

On page 173 of this issue, Soutschek *et al.*¹ describe a simple but effective method for the intravenous delivery of nucleic acids called short interfering RNAs that target a therapeutically important messenger RNA molecule. Messenger RNAs (mRNAs) represent a necessary step in gene expression, being the intermediate between gene and protein. Short interfering RNAs (siRNAs) are duplexes of RNA, 21 to 23 nucleotides long, one of whose strands becomes incorporated into a complex of proteins. This small RNA component serves as a guide to identify a complementary sequence in an mRNA, either causing the mRNA to be cleaved by the protein complex (Fig. 1), or preventing it from being translated into protein^{2–5}. The result is that the encoding gene is silenced. It is hoped that this approach might be used to shut down disease-related genes in humans; with Soutschek and colleagues' findings in mice, that dream moves a little closer to reality.

It was first demonstrated in the late 1970s that the specificity of bonding between

complementary nucleic-acid strands could be exploited to create specific inhibitors of gene expression⁶. This heralded the advent of what researchers now call 'antisense-mediated inhibition of gene expression'. Early work in the field used short, chemically synthesized strands of DNA that could be designed to selectively bind, through complementary base pairing, to any target mRNA of interest, and direct its cleavage through a cellular enzyme called ribonuclease H.

This selectivity of bonding between two nucleic-acid strands led to the idea that such antisense mechanisms might provide highly selective drugs that specifically silence harmful genes, of both viral and non-viral origin. The idea received a huge boost when it was discovered that a natural antisense mechanism, called RNA interference (RNAi)⁷, exists in nearly all organisms except bacteria. In RNAi, siRNAs are produced naturally by enzymatic cleavage from longer, double-stranded precursor molecules. They then act as guides for the cleavage (or prevention of translation) of target mRNAs by a protein

complex known as the RNA-induced silencing complex (RISC; Fig. 1). Excitement over the potential applications of RNAi was fuelled by the finding that chemically synthesized siRNAs could be delivered, in complex with lipids, to human cells in culture, triggering sequence-specific silencing of complementary mRNAs^{8–10}.

The scientific and investment communities immediately seized on these findings, realizing that this powerful cellular mechanism might be harnessed to create designer drugs, and would perhaps surpass conventional DNA-based antisense technologies in both application and market value. Yet there was a problem: although the potency of siRNAs in cultured cells was beyond doubt, it was difficult to deliver these duplex RNAs effectively to tissues *in vivo*. Small synthetic DNA molecules, with selective chemical modifications to their sugar-phosphate backbones, can be delivered to tissues and cells *in vivo* by simple intravenous injection. But siRNAs can tolerate only a few modifications and are not readily taken up into tissues. They are also susceptible to degradation by enzymes in the blood known as nucleases.

Soutschek *et al.*¹ now present a relatively simple breakthrough in this problem of delivery. The group decided to target the mRNA that encodes apolipoprotein B, a molecule involved in the metabolism of cholesterol. The concentrations of this protein in human blood samples correlate with those of cholesterol, and higher levels of both compounds are associated with an increased risk of coronary heart disease.

Soutschek *et al.* synthesized a series of siRNAs targeting the apolipoprotein B mRNA; these siRNAs contain selective stabilizing modifications and are joined to a cholesterol group that is chemically linked to the terminal hydroxyl group of the sense-strand RNA. Intravenous injections of the siRNA-cholesterol conjugates in mice resulted in uptake into several tissues, including the liver, jejunum (part of the small intestine), heart, kidneys, lungs and fat tissue. Most importantly, the siRNAs efficiently reduced the levels of apolipoprotein B mRNA by more than 50% in the liver and by 70% in the jejunum. This reduction resulted in a lowering of the levels of blood cholesterol comparable to that observed in mice in which the apolipoprotein B gene had been deleted.

The beauty of these results is the relative

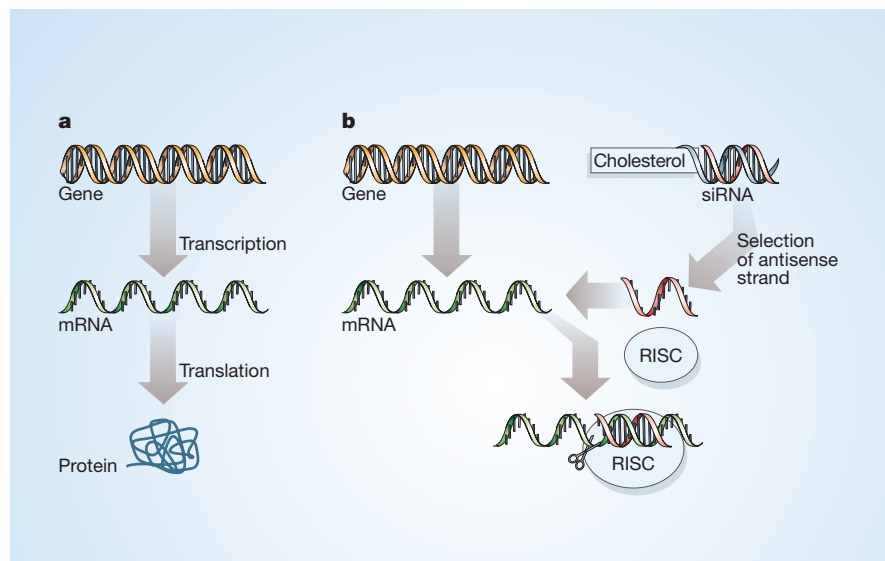


Figure 1 Silencing genes the RNAi way. **a**, For a gene to be expressed, its DNA sequence must be copied (transcribed) into messenger RNA (mRNA); this must in turn be translated into a protein sequence. **b**, RNAi works by either destroying the mRNA (bottom) or preventing it from being translated (not shown). In Soutschek and colleagues' modification¹ of the general RNAi approach, short interfering RNAs (siRNAs) are synthesized, chemically modified and labelled on the 'sense' strand (blue) with cholesterol. The siRNAs are then injected intravenously into mice, where the cholesterol group enables the siRNAs to be taken up into tissues. There, the sense strand is destroyed by the inherent RNAi pathway, leaving the antisense strand (red) to bind to a complementary sequence in a target mRNA. Recruitment of a protein complex, the RNA-induced silencing complex (RISC), enables the mRNA to be cleaved.

simplicity of the delivery method. The system did not require expensive lipid complexes or other macromolecular carriers, but merely a single cholesterol conjugate per RNA duplex. That cholesterol group also provides the key to the authors' success. Without cholesterol, there was no tissue delivery of the chemically modified siRNAs, despite their resistance to blood nuclease enzymes. The use of cholesterol helped cells to take up the siRNAs.

To verify that the siRNAs were functioning through RISC-mediated mRNA cleavage, Soutschek *et al.* also identified the point at which the mRNA is severed. They found that cleavage occurred at a position corresponding to ten nucleotides from one end (the 5' end) of the siRNA — precisely where RISC cleavage is known to occur¹¹. Moreover, the cleavage products were observed only in RNAs obtained from animals treated with functional siRNA-cholesterol conjugates.

So Soutschek and colleagues' results are most encouraging. Yet many questions must be addressed before the method sees application in humans. For instance, the treatment of high cholesterol levels in humans might require lifetime use of cholesterol-lowering compounds. But the consequences of long-term siRNA therapy are not known, nor is it known whether the benefit would exceed any risk. Another potential problem is the dosage required to elicit the desired effect. Translating the mouse data to humans, gram quantities of siRNA-cholesterol conjugates would be required in regular infusions, which would be enormously expensive.

These concerns notwithstanding, it is remarkable that, only a few years after the discovery and description of the mechanism of RNAi, a potential method for disease treatment has been reported. By comparison, it was well over a decade before antisense DNA showed similar promise. It remains to be seen whether the cholesterol-conjugate approach can be used to silence other disease-related genes in animal models. If so, it should revolutionize the use of RNAi. ■

John J. Rossi is in the Division of Molecular Biology, Graduate School of Biological Sciences, Beckman Research Institute of the City of Hope, Duarte, California 91010, USA.

e-mail: jrossi@bricoh.edu

1. Soutschek, J. *et al.* **432**, 173–178 (2004).
2. Novina, C. D. & Sharp, P. A. *Nature* **430**, 161–164 (2004).
3. Mello, C. C. & Conte, D. Jr *Nature* **431**, 338–342 (2004).
4. Hannon, G. J. & Rossi, J. J. *Nature* **431**, 371–378 (2004).
5. Tuschl, T. & Borkhardt, A. *Mol. Interv.* **2**, 158–167 (2002).
6. Zamecnik, P. C. & Stephenson, M. L. *Proc. Natl Acad. Sci. USA* **75**, 280–284 (1978).
7. Fire, A. *et al.* *Nature* **391**, 806–811 (1998).
8. Elbashir, S. M. *et al.* *Nature* **411**, 494–498 (2001).
9. Elbashir, S. M., Lendeckel, W. & Tuschl, T. *Genes Dev.* **15**, 188–200 (2001).
10. Caplen, N. J., Parrish, S., Imani, F., Fire, A. & Morgan, R. A. *Proc. Natl Acad. Sci. USA* **98**, 9742–9747 (2001).
11. Schwarz, D. S., Tomari, Y. & Zamore, P. D. *Curr. Biol.* **14**, 787–791 (2004).

Earth science

Mantle mapped in the desert

Georges Ceuleneer

How does variation in ocean-floor rocks arise from differences in the temperature of their mantle source? A new angle on the question comes from painstaking work on one of the geological wonders of the world.

One hundred million years ago, as the Tethys Sea separating the supercontinents of Gondwanaland and Laurasia closed, a huge chunk of mid-oceanic rocks was pushed up and onto the neighbouring continent. These rocks are now exposed in the mountains of northern Oman, running from the sands of Arabia to the Indian Ocean, and they constitute the 'Oman ophiolite' — at 30,000 km², possibly the largest outcrop of ocean crust and upper mantle exposed on the continents and one of the geological wonders of the world. This is not any old ocean floor, but a 'fossil' spreading centre that was once the site of undersea volcanic activity.

It is for these reasons that the Oman ophiolite is one of the best-studied geological bodies in the world. A report of the latest investigations appears on page 167 of this issue¹, where Le Mée *et al.* describe field and lab analyses aimed at understanding how events in the mantle underlying oceanic spreading centres produce ridge segmentation. Spreading centres are essentially linear features, but they are characterized by frequent offsets ranging in size from a few kilometres to hundreds of kilometres. The origin of this segmentation is still debated.

The broad context for these studies is that the face of the Earth is being continually remodelled by tectonic activity, volcanism and erosion. Underlying all of these processes is solid-state convection in the mantle — the 'rocky' shell that envelops the metallic core, and extends from the base of the thin crust on which we live to a depth of 2,900 km. At the temperatures of more than 1,000 °C that are typical of Earth's interior, the silicate minerals that make up the mantle can flow at rates of several centimetres per year. Mantle convection is driven by contrasts in density induced by the thermal expansion of silicates (the hotter a rock, the lighter it is).

So determining the thermal structure of the mantle is a major objective in Earth science. At present we have no way of directly measuring the temperature of Earth's interior at mantle depths, and instead have to use indirect methods. One such approach, and probably the least indirect, is to analyse the chemical composition of basalts — rocks that solidified from magmas erupted at the surface at places such as spreading centres. In this peculiar setting, basaltic melts are generated by the partial melting of mantle rocks induced by decompression during mantle

upwelling. The composition of these melts is a function of the degree of partial melting: melts produced by small amounts of melting are relatively enriched in the so-called 'incompatible elements', such as sodium, titanium and lanthanum, that preferentially enter the liquid during melting. The relations between the mantle temperature, the amount of decompression and the degree of partial melting, together with the solid-liquid partition coefficient for many chemical elements, have been calibrated experimentally, making it theoretically possible to calculate the temperature of the mantle below a given site of volcanic activity from the chemical composition of erupted basalts.

In practice, 'inversion' of chemical composition to calculate mantle temperature requires assumptions as to whether the mantle source is chemically homogeneous or not; how melt fractions produced over a large region aggregate to produce basalt; and the extent to which the composition of the basalt has been modified during its ascent through the mantle and crust. In other words, there is no consensus concerning the reliability of models relating basalt composition to mantle temperature.

In the new work, Le Mée *et al.*¹ provide fresh evidence supporting the idea, initially proposed by Klein and Langmuir², that part of the variability in the chemical composition of basalts erupted at oceanic spreading centres can be related to the thermal structure of the mantle. The variable composition of such basalts is well documented: tens of thousands of samples, collected worldwide, have been analysed. An ideal way to test Klein and Langmuir's hypothesis would be to establish, with the same resolution as for basalt, a map showing how the composition of the mantle residue varied as a result of partial melting. This exercise cannot be conducted easily in present-day oceans, where mantle rocks occur only patchily, in settings where the basaltic crust is anomalously thin, and where they are usually greatly altered by interaction with sea water.

Hence Le Mée and colleagues' interest in the Oman ophiolite, a geological centre of attention since the work of pioneers such as Bob Coleman and Tjerk Peters^{3,4}. Here, rocks from the ocean crust, including basalts, and the underlying mantle rocks are exposed; and in their chemical composition they retain a record of melting and crystallization processes that are similar to those occurring



Figure 1 Rough country. The Oman ophiolite, one of the largest outcrops of ocean crust and upper mantle exposed on the continents, forms a 400-km-long part of the mountains of northern Oman.

along present-day oceanic spreading centres. But no previous geochemical analysis of the Oman mantle rocks has been based on such a detailed field survey as that conducted by Le Mée and colleagues. Their high-resolution sampling along the 400 km of these rugged mountains took many months; Figure 1 shows the terrain they had to work in. This 'exploration geology' approach, with the ensuing analytical data firmly anchored in the context of field geology, is, alas, less and less in vogue in the academic community.

Back in the laboratory, Le Mée *et al.* found that the chemical composition of Oman mantle rocks varies considerably between samples. In particular, incompatible elements show significant variation: simple mass-balance considerations were used to calculate that the least-depleted samples correspond to a relatively low degree of partial melting (about 10%), whereas the more-depleted samples correspond to partial melting reaching 30%. This range is similar to that estimated from the composition of ocean-floor basalts².

Le Mée and colleagues' most notable result is related to the geographical distribution of their data. They show that the variations in composition of Oman mantle rocks are not random but have 'highs and lows', separated by tens of kilometres, allowing the authors to define a chemical segmentation of this fossil spreading centre. This pattern is reminiscent of those defined on a morphological and geochemical basis along certain segments of present-day oceanic spreading centres. Translated into mantle temperatures — not, as I've said, a straightforward exercise — these variations in chemical composition could correspond to temperature contrasts of several tens of degrees, which are high enough to drive mantle convection. In that respect, Le Mée *et al.* could have

provided the missing link allowing us to relate segmentation of mid-ocean ridges to small-scale convection processes in the shallow mantle⁵.

One possible confounding factor, however, is that a variable degree of partial melting is not the only way to account for the variable composition of residual mantle rocks. The shallow mantle may act as a 'reactive filter' whose composition, particularly

in the concentrations of incompatible elements, may be permanently modified by complex reactions with magmas percolating to the surface⁶. Le Mée *et al.* have made every effort to sample areas apparently not affected by such mantle–melt reactions. But we cannot be entirely sure that they have avoided the cryptic effects of these reactions — effects that have changed mineral composition but not mineral proportions.

It is often the case that later studies fail to support this or that scientific model of a natural process, and such could be the fate of Le Mée and colleagues' interpretation of chemical variability at spreading centres in terms of partial melting. But the map they have patiently extracted from the barren desert rocks of Oman will endure: it will prove an invaluable document in debates about basalt composition, mantle processes and the thermal structure of the Earth. ■

Georges Ceuleneer is at the Observatoire Midi-Pyrénées, CNRS, 14 Avenue E. Belin, 31400 Toulouse, France.

e-mail: georges.ceuleneer@cnes.fr

1. Le Mée, L., Girardeau, J. & Monnier, C. *Nature* **432**, 167–172 (2004).
2. Klein, E. M. & Langmuir, C. H. *J. Geophys. Res.* **92**, 8089–8115 (1987).
3. Coleman, R. G. *J. Geophys. Res.* **86**, 2497–2508 (1981).
4. Allemann, F. & Peters, Tj. *Ecolgne Geol. Helv.* **65**, 657–697 (1972).
5. Briais, A. & Rabinowicz, M. *J. Geophys. Res.* **107**, ECV-3, 1–17 (2002).
6. Seyler, M., Toplis, M. J., Lorand, J.-P., Luguét, A. & Cannat, M. *Geology* **29**, 155–158 (2001).

DNA repair

Big engine finds small breaks

Anna Marie Pyle

When a break occurs in the DNA double helix, it must be dealt with rapidly. The structure of one of the cellular machines responsible is now revealed, offering insights into its impressive speed and flexibility.

A double-strand break in DNA is the cellular equivalent of a three-alarm fire. As any taxpaying citizen knows, most communities maintain an impressive collection of hardware for dealing with major emergencies. Cells are no different, and for bacteria, the RecBCD complex of proteins is the heavy machinery that rushes to the scene of the conflagration¹. On page 187 of this issue, Wigley and colleagues² unveil the crystal structure of *Escherichia coli* RecBCD interacting with a DNA break. This remarkable structure helps us to visualize how the complex careers along the DNA, applies the brakes at a designated site, and chops up unwound DNA in its wake.

RecBCD aids in the repair of double-strand breaks that occur during normal DNA replication and as a result of damage by ionizing radiation¹. It is a tight complex of three proteins, two of which, RecB and RecD, have

the ability to unwind DNA (they are helicases). These two motors move in the same direction along a DNA double helix, but on opposite strands (hence they move with opposite polarity, RecB progressing in the so-called 3' to 5' direction, and RecD in the 5' to 3' direction). In this way, they can push together against a duplex section of DNA, opening it^{3,4} (Fig. 1, overleaf). RecB also chops up the unwound DNA. RecC, meanwhile, has been implicated in the recognition of 'signal' sequences in DNA, known as Chi sequences (5'-GCTGGTGG-3'), that direct the complex to slow down⁵. Remarkable features of RecBCD include its unprecedented speed and processivity — its ability to remain associated with its substrate — together with a puzzling affinity for DNA that is terminated by a clean cut (a 'blunt end')¹.

The RecBCD structure now published by Wigley and colleagues² reveals an elaborate

complex of interlocked proteins that have invaded a DNA duplex, opened four base pairs of DNA and sent the separated single strands in opposite directions. The 3'-terminated strand is fed towards the RecB motor and the 5'-terminated strand is fed towards the RecD motor. RecC appears to manage this operation, holding the helicase domains of RecB and RecD in place, splitting the DNA and directing the orientation of the divided strands.

The structure is particularly significant because it provides a molecular framework for understanding the recognition of Chi DNA by RecC, a protein that has long been mysterious because it shares little sequence similarity with known protein families. Wigley and co-workers show that RecC is no orphan: it has the architecture of an SF1-type helicase that has lost key catalytic amino acids. Indeed, the 3'-terminated strand of DNA passes over a region of RecC that is typically used as a helicase–nucleic-acid interface. The structure suggests a remarkable model in which RecC has exchanged the ability to actively unwind DNA for the ability to scan it and then apply the brakes when it passes over a Chi sequence.

Meanwhile, the structural features of RecB help to explain why its 'nuclease' domain occasionally interrupts its continuous chewing of the 3'-terminated strand of the DNA (its exonuclease activity) in order to take a bite out of the other strand (endonuclease activity), particularly when the complex has paused at a Chi site. RecB's helicase and nuclease domains are connected by a long tether that provides sufficient freedom for the nuclease to take occasional swipes at the 5'-terminated strand.

Although this work provides ample insight into the molecular mechanisms for repairing DNA, it also contains rewards for simple 'motorheads' such as myself who want to understand how proteins move along, and work on, DNA and RNA. For example, it goes a long way towards answering a basic question: how can a helicase move so fast? RecBCD unwinds DNA at around 1,000 base pairs per second, which is the fastest machinery yet discovered for unwinding nucleic-acid duplexes. And yet the structure shows that the RecB and RecD motors look like generic SF1 helicases, which typically mosey along their substrates with comparatively unexceptional velocities^{6,7}. Indeed, RecB and RecD unwind DNA slowly in isolation.

Although there is probably synergy in the coupled motion of RecB and RecD, a key to the speed of the complex may lie with the component that is not a motor: RecC. By surrounding and splitting the DNA strands before feeding them to the motors, RecC might reduce the tendency of the divided strands to clamp shut, allowing RecB and RecD to move faster. RecC is also likely to enhance processivity by effectively tying

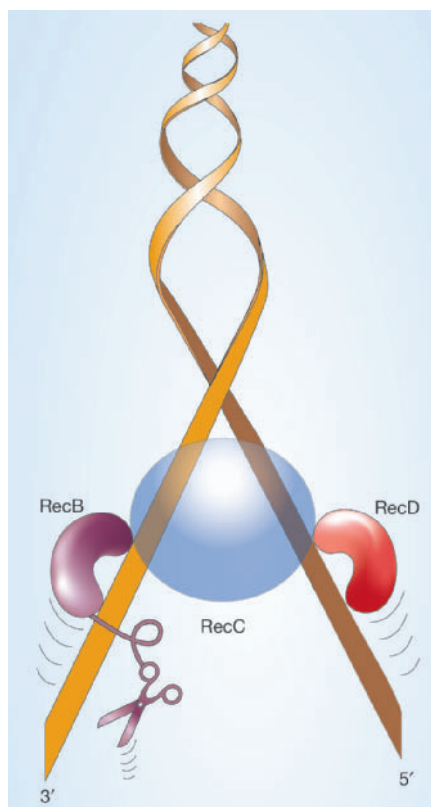


Figure 1 Heavy machinery for dealing with DNA breaks. The RecBCD complex of proteins helps to repair DNA by unwinding the DNA double helix in the region of a double-stranded break, and chopping up the unwound strands. RecB and RecD are helicases that push on the DNA duplex together to unwind it. RecB also has 'nuclease' activity, which snips up the unwound DNA. RecC has been implicated in recognizing DNA sequences that direct the complex to slow down. Wigley and colleagues² have now presented the crystal structure of RecBCD bound to DNA, revealing details such as how the complex binds DNA at the site of a break and how the actions of the three proteins are coordinated.

RecB and RecD to the substrate. Whatever its mechanism of action, RecC clearly engages the DNA before delivering it to the motors, thereby turbo-charging two otherwise unexceptional engines.

The structure also helps us understand how a helicase can move backwards. Most SF1 and SF2 helicases move along single strands of DNA or RNA in a 3' to 5' manner. However, single-molecule studies have detected apparent backward motion by helicases⁸, and there are examples of closely related helicases (such as Dda and, of course, RecD) that move in the opposite direction^{3,4,9}. The new structure shows that RecD does not do this by turning around—that is, the helicase domains are not arranged in an opposite orientation relative to the polarity of the DNA strand. Rather, RecD simply drives in reverse, much like variants of the cytoskeletal motors myosin and kinesin¹⁰.

This structural evidence for forward and reverse gears in a helicase opens the door to structure–function studies aimed at understanding directionality of motion.

Insights into how helicases and substrates recognize each other are also provided. Most helicases require a single-stranded launch pad attached to the duplex before they can initiate unwinding¹¹. But RecBCD can load directly onto a DNA blunt end, and the structure shows how: it creates its own single-stranded tails by melting the first four base pairs of the duplex in the absence of adenosine triphosphate (ATP, the fuel for biological motors such as helicases). So the intrinsic binding energy of the motor makes major contributions to unwinding events prior to any conformational changes that are induced by ATP hydrolysis.

Despite the appealing mechanical concepts suggested by the new structure, however, it should be emphasized that it represents a snapshot of an initiation complex that has not undergone forward motion in the presence of ATP. The actively translocating complex may, in fact, contain features that differ markedly from those seen here. For example, a compelling feature of the complex is the 'pin' provided by RecC that appears to split the DNA strands before they are delivered to RecB and RecD. However, the pin substructure may be subjected to substantial force during forward motion. It will be interesting to see whether it persists during unwinding, and to understand its mechanical role in RecBCD complexes that are actively moving.

Similarly, it will be essential to explore the existence of the proposed binding site between RecC and the Chi sequence, and to test the suggested model by which the Chi–RecC interaction regulates the nuclease activity of RecB. Like any significant crystal structure, that of RecBCD poses intriguing new questions. In this case, it also provides a valuable roadmap for mutational and mechanistic analyses.

Anna Marie Pyle is in the Department of Molecular Biophysics and Biochemistry, and the Howard Hughes Medical Institute, Yale University, 266 Whitney Avenue, New Haven, Connecticut 06520, USA.

e-mail: anna.pyle@yale.edu

1. Kowalczykowski, S. C. *Trends Biochem. Sci.* **25**, 156–165 (2000).
2. Singleton, M. R., Dillingham, M. S., Gaudier, M., Kowalczykowski, S. C. & Wigley, D. B. *Nature* **432**, 187–193 (2004).
3. Dillingham, M. S., Spies, M. & Kowalczykowski, S. C. *Nature* **423**, 893–897 (2003).
4. Taylor, A. F. & Smith, G. R. *Nature* **423**, 889–893 (2003).
5. Handa, N., Ohashi, S., Kusano, K. & Kobayashi, I. *Genes Cells* **2**, 525–536 (1997).
6. Cheng, W., Hsieh, J., Brenda, K. M. & Lohman, T. M. *J. Mol. Biol.* **310**, 327–350 (2001).
7. Maluf, N. K., Fischer, C. J. & Lohman, T. M. *J. Mol. Biol.* **325**, 913–935 (2003).
8. Dessinges, M. N., Lionnet, T., Xi, X. G., Bensimon, D. & Croquette, V. *Proc. Natl Acad. Sci. USA* **101**, 6439–6444 (2004).
9. Raney, K. D. & Benkovic, S. J. *J. Biol. Chem.* **270**, 22236–22242 (1995).
10. Vale, R. D. & Milligan, R. A. *Science* **288**, 88–95 (2000).
11. Delagoutte, E. & von Hippel, P. H. *Q. Rev. Biophys.* **35**, 431–478 (2002).

Ecology

Hope in the hills for tundra?

Peter D. Moore

Will global warming cause northern forests to spread into arctic tundra? A study of black spruce suggests that the answer is complex and varies according to latitude and altitude.

Global warming is likely to spell trouble for the tundra. The polar tundra has nowhere to retreat to when boreal forests invade from warmer latitudes, and alpine tundra has no alternative but extinction when tree lines climb to the mountain tops. These pessimistic models assume that forest spread into tundra habitats is inevitable in a warmer world, but that assumption needs to be questioned and tested. Isabelle Gamache and Serge Payette¹, writing in the *Journal of Ecology*, have now examined evidence for changes in tree growth in the zone of contact between tundra and forest in the east of Canada. They have found that the expansion of forest is not entirely following the expected pattern.

Climate modelling indicates that the Arctic will experience an even more rapid rise in temperature than most other parts of the globe². It is therefore reasonable to suppose that the trees currently occupying regions immediately south of the tundra will respond to the warming by increasing their individual growth, their reproduction, and their invasive capacity. Gamache and Payette¹ have paid particular attention to growth processes, because they regard vegetative growth as a fundamental prerequisite for the reproduction and consequent spread of trees. The black spruce (*Picea mariana*), which is the main tree found in the zone between boreal forest and arctic tundra in eastern Canada, grows in a stunted, dwarf form at the northern limit of its distribution and at higher altitudes, creating a vegetation known as krummholz (Fig. 1). Trees with this growth form have a very low reproductive capacity, so increased vegetative growth might be expected to be a prelude to increased seed production and extension of the range of the tree.

Gamache and Payette sampled black spruce trees from tree-line locations along a 300-km transect ranging from 55° to 58° N, extending from open boreal forest in the south to shrub tundra with fragmented krummholz patches in the north. Patterns of growth over time were studied by examining the height of trees in relation to age (determined by annual growth rings), and also by measuring the annual extension of leader shoots (the upper parts of vertical stems) in recent years.

Their measurements reveal that the northern populations have been suppressed in their growth in the past, but in the last ten

years or so their growth has accelerated and is now becoming comparable to that of the southern forest trees. This growth, however, has not yet been accompanied by increased regeneration by seed. In contrast, when trees at higher altitude, scattered above the tree line, were examined, there was little evidence of increased growth, particularly in the more southerly alpine sites. Here, exposure to wind, the speed of which is greater above the tree line, seems to be holding back any surge in spruce growth as a response to higher summer temperatures. Unlike at northern sites, seed regeneration does occur, but further growth is suppressed.

Both historical³ and physiological⁴ studies have shown that the responses of the tundra-edge forests to climate change are complex, and the findings of Gamache and Payette bear this out. The influence of snowfall is an important consideration and this seems to have increased in many subarctic regions over the past century⁵. More snow can result in later spring melting and hence a shorter growing season for the trees, which



Figure 1 Transition zone: krummholz vegetation, a stunted form of black spruce trees. Gamache and Payette¹ have investigated the growth patterns of black spruce in eastern Canada.

will hold back any northward advance of the forest. Indeed, Gamache and Payette found that season length (measured as the number of frost-free days in which the temperature exceeded the minimum required for tree growth) was the major constraint on leader-shoot extension in black spruce. Microclimate, determined by slope and aspect, will also have a significant role in any boundary changes for the forest; for instance, in regions where survival is marginal, there will still be locally favourable sites for tree growth⁶.

Much will hang upon whether the enhanced growth of the krummholz spruce vegetation¹ will later result in increased seed production and consequent further invasion of the shrub tundra to the north. On the southerly mountain tops, meanwhile, there is an abundant supply of wind-borne seeds that can now germinate in the exposed but unshaded alpine tundra. As the new findings suggest, however, it is possible that wind exposure will prevent these seedlings from growing into tall trees, allowing the development merely of krummholz vegetation. It remains to be seen whether this montane wind-stunting effect is geographically widespread, for there is evidence of upward tree-line movements on high-latitude mountains in Scandinavia⁷, and tree lines globally seem to follow a common temperature threshold⁸.

The ecological implications of studies such as this are considerable. The increased biomass represented by faster tree growth in the Arctic¹ could be viewed as a valuable additional sink for atmospheric carbon⁹. In contrast, if there were additional tree growth on the high-latitude wetlands, this could lead to increased evaporation and transpiration, and hence to surface drying. This, in turn, would stimulate soil respiration in the drying peatlands and could possibly enhance methane production, adding to the atmospheric load of greenhouse gases.

Conservationists have also expressed concern that any forest expansion in the tundra could present a risk to the survival of arctic-alpine plants and animals. The tundra may not be very biodiverse, but it contains a distinctive assemblage of organisms that require conservation, especially if threatened by forest invasion. There is, however, a resilience among arctic ecosystems¹⁰, as shown by their survival through the extreme climatic fluctuations of the Pleistocene era (roughly the past 2 million years)¹¹. The flora and fauna of the tundra probably evolved originally in a montane setting, and the high-latitude mountains, according to the findings of Gamache and Payette, may well prove to be their best hope for refuge in the coming heat.

Peter D. Moore is in the Division of Life Sciences, King's College London, Franklin-Wilkins Building, 150 Stamford Street, London SE1 9NN, UK. e-mail: peter.moore@kcl.ac.uk

1. Gamache, I. & Payette, S. *J. Ecol.* **92**, 835–845 (2004).

W. P. CONWAY/CORBIS



100 YEARS AGO

The columns of daily papers have during the last two weeks contained many references to the question of the origin of life... Prof. Ray Lankester and Dr. Chalmers Mitchell... proclaim themselves, as followers of Huxley, believers in evolution generally, and in the natural origin of living matter in the past. They, like many others, refuse to believe that it takes place at the present time, because undoubted proof of its occurrence cannot be produced by laboratory experiments. The uniformity of natural phenomena would certainly lead us to believe, as Sir Oliver Lodge has intimated, that if such a process occurred in the past, it should have been continually occurring ever since — so long as there is no evidence to show cause for a break in the great law of Continuity. Certainly no such evidence has ever been produced, and if the origin of living matter takes place by the generation in suitable fluids of the minutest particles gradually appearing from the region of the invisible, such a process may be occurring everywhere in nature's laboratories, though altogether beyond the ken of man. From *Nature* 10 November 1904.

50 YEARS AGO

Prof. Max Born, who has been awarded a share of the Nobel Prize in Physics for 1954, is known for many contributions to modern theoretical physics, particularly to the development of quantum mechanics and to the theory of crystals. The work mentioned specifically in the announcement of the Nobel award is the interpretation of the wave functions as probabilities for the positions of particles, a vital step in the development of the modern view of the relations between particle and wave aspects of atomic theory.

Prof. W. Bothe, of the University of Heidelberg, who shares the 1954 Nobel Prize in Physics with Prof. Born, is known for many important contributions to modern physics. The best known of these include his introduction of coincidence methods into counting techniques and the work in which, together with Geiger, he applied the coincidence method to the Compton effect and showed that the conservation laws are satisfied in each individual event and not merely on the average. This was fundamental for the interpretation of atomic processes. From *Nature* 13 November 1954.

2. Houghton, J. T. et al. (eds) in *Climate Change 2001: The Science of Climate Change* (Cambridge Univ. Press, 2001).
3. Payette, S., Filion, L., Delwaide, A. & Begin, C. *Nature* **341**, 429–432 (1989).
4. Jarvis, P. & Linder, S. *Nature* **405**, 904–905 (2000).
5. Vaganov, E. A., Hughes, M. K., Kiryanov, A. V., Schweingruber, F. H. & Silkin, P. P. *Nature* **400**, 149–151 (1999).
6. Lennon, J. J., Kunin, W. E., Corne, S., Carver, S. & Van Hees, W. W. S. *Glob. Ecol. Biogeogr.* **11**, 103–114 (2002).
7. Kullman, L. J. *Ecol.* **90**, 68–77 (2002).
8. Körner, Ch. & Paulsen, J. J. *Biogeogr.* **31**, 713–732 (2004).
9. Loya, W. M. & Grogan, P. *Nature* **431**, 406–408 (2004).
10. Wookey, P. A. & Robinson, C. H. *Opera Botanica* **132**, 215–232 (1997).
11. Crawford, R. M. M. & Abbott, R. J. *Botanica Acta* **107**, 271–278 (1994).

Animal behaviour

Rank crime and punishment

Joan E. Strassmann

In paper wasps, facial markings are cheap 'status badges' that would seem to be susceptible to cheating. But wasps punish those whose markings lie. Social competition is, it appears, a strong selective force.

'Badges of status' are markings that animals are thought to use to signal their size and dominance — they are indicators of rank. To be useful, however, they must be 'honest' indicators; a symbol of high rank cannot be adopted by a low-quality individual. How can such cheating be prevented? The idea that badges of status must necessarily be honest because they are simply too costly for inferior individuals to produce has gained little experimental support. But there is a theory that social costs might instead be imposed on cheaters. On page 218 of this issue, Tibbetts and Dale¹ provide support for this hypothesis. They show that in female paper wasps (*Polistes dominulus*), highly variable facial marks are related to overall size, and are used as rank indicators. Moreover, individuals painted with a mark that indicates a higher rank than they came by naturally are punished — something that has been difficult to demonstrate in other systems.

A century ago, Wilhelmine Enteman² painstakingly documented within-colony colour variation in the social wasp *Polistes*. The purpose of such variation was unclear — communication was a possible function, but this was not investigated at the time. Subsequent work on communication among social insects instead focused on variation in nest and wasp chemicals. In wasps it was found, for instance, that kin recognition is based on variations in non-volatile, long-chain hydrocarbons found on the insects' surface³. Recently, however, Tibbetts⁴ re-examined the visual colour marks and found that they are important in communication after all.

Social wasps need to be able both to distinguish between individuals and to recognize the individuals' status, in order to establish and maintain a hierarchy of dominance. Mated females begin new nests together, and the winner of dominance contests lays most of the eggs. This dominance is usually expressed in low-cost ritualized postures rather than serious fighting⁵.

Tibbetts⁴ found that *Polistes fuscatus* females use individual identity cues based on facial and abdominal marks to recognize others in the hierarchy. She also showed that females become aggressive towards known females with altered marks.

Now, Tibbetts and Dale¹ have examined the markings on *P. dominulus* females to see whether they are used as status badges. A status badge is a detectable indicator of rank that is related to a measure of competitive ability^{6,7}. Compared with individual identity cues, status badges are likely to be less variable, more dependent on physical condition and less strongly genetically determined⁸. The idea of badges of status was first invoked to explain rank-correlated variation in colour patterns among feeding flocks of birds that have frequent dominance interactions involving greater numbers of individuals than the birds could easily learn to identify individually⁶. A status badge would allow unfamiliar individuals to quickly determine, without aggression, who should give way. The concept has also been applied to territorial contests in which predicting the likelihood of losing would be advantageous⁹.

Polistes dominulus has highly variable dark spots on the clypeus (the region just above the mouth; Fig. 1). As Tibbetts and Dale describe¹, spot number and brokenness — a measure that includes the number and irregularity of the spots — are slightly greater in wasps with wider heads. Head width is a well-known correlate of overall size and dominance in wasps. Thus the authors found, as predicted, that females with more spots or a higher brokenness index were more likely to become dominant in two-player contests.

A major question about colour-based status badges is why they do not seem to be vulnerable to cheating, even though they would appear to be 'cheap' to make. One hypothesis invokes social costs: an impostor would have to engage in more dominance contests, but would be likely to lose because of poor competitive ability. Support for this

hypothesis has been equivocal in other systems, but now appears clear in *P. dominulus*. Losers were more likely to suffer continuing aggression from the dominant female if they had a more broken (higher-rank) pattern.

Going beyond correlation, Tibbetts and Dale manipulated the clypeus spots with yellow and black paint. To one member of a pair that had not previously interacted, the authors either added a spot, obliterated an existing spot, or added paint in ways that did not change the existing pattern. After these females had established a clear dominance relationship, dominant females continued to be aggressive towards subordinates whose painted marks indicated an increase in status. The result clearly supports the hypothesis that there are social costs to erroneously high status signals: cheating subordinates are punished.

The study had several puzzling features, however. The key result discussed above relates to behaviour after dominance had been established. Surprisingly, though, manipulating the spot pattern had no effect on behaviour before rank was established. Also, females painted with high-status marks were no more likely to attain dominant status than were those painted with low-status marks, or controls. If extra marks give a badge of high status, then their owner should be more likely to become dominant.

Another curiosity is that after dominance was established, the dominant females were also more aggressive towards subordinates painted to look even more subordinate. This is hard to explain in a two-female system. But wasps often begin nests in groups of more than two. A female with marks indicating very low rank that acted instead like a second-ranked female could plausibly be perceived by the dominant female as a threat and receive more aggression. A further unpredicted result is that unpainted subordinates were aggressive towards painted dominants, whether they were painted with a higher- or a lower-status mark than that of their 'true' status. In ten cases, the unpainted subordinate female even overthrew the painted dominant; these cases were evenly divided between those in which the dominant was painted with a higher- or a lower-status mark.

Also interesting is that the correlation of spot number or brokenness with head width is very weak, accounting for only 7% or 3% of the variance, respectively. A reliable cue of condition should account for more of the variance. Perhaps head width is not the best measure of condition, and perhaps the clypeus marks are actually correlated with a better, as-yet-unknown measure. But the marks are fixed in the pupal stage, and are not altered by later feeding and overwintering, which are known to affect condition. It seems that if clypeus marks are status badges, they are not conventional ones.



Figure 1 Badge of status. *Polistes dominulus* wasps show variation in the number and brokenness of spots on the clypeus, above the mouth. Tibbetts and Dale¹ find that these measures provide an indicator of how dominant or subordinate an individual is; dishonesty brings social costs.

Tibbetts and Dale were careful to control for numerous factors that might have influenced the outcome of the contests. First, they paired wasps for mass. Second, they paired individuals that were unrelated and came from sites kilometres apart, so they could not have had any prior interactions. (In this species, females begin nests with relatives or nearby non-relatives with whom they are likely to have previously interacted⁹.) A further reason why the paired individuals could not have benefited from recognizing each other is that they were videotaped immediately after pairing. However, it is possible that they normally use individual recognition as well as status badges, as individual recognition can occur, at least in another *Polistes* species⁴.

It might seem remarkable that wasps have such sophisticated abilities. But social competition is a very strong selective force¹⁰. Pinyon jays can infer rank by watching others interact¹¹. Viruses can cheat¹². Perhaps it should be no surprise that social wasps can recognize individuals and their rank.

Joan E. Strassmann is in the Department of Ecology and Evolutionary Biology, Rice University, Houston, Texas 77251-1892, USA.

e-mail: strassm@rice.edu

1. Tibbetts, E. A. & Dale, J. *Nature* **432**, 218–222 (2004).
2. Enteman, W. M. *Carnegie Inst. Wash. Publ.* **19**, 1–88 (1904).
3. Singer, T. L., Espelie, K. E. & Gamboa, G. J. in *Pheromone Communication in Social Insects: Ants, Wasps, Bees, and Termites* (eds Vander Meer, R. K., Breed, M. D., Espelie, K. E. & Winston, M. L.) 104–125 (Westview, Boulder, CO, 1998).
4. Tibbetts, E. A. *Proc. R. Soc. Lond. B* **269**, 1423–1428 (2002).

Plant development

A bunch of leaves

The intricate and elegant architecture of a flower requires the activity of a plethora of proteins. Many belong to a group of gene regulators known as MADS-box proteins. These form multiprotein complexes that direct the formation of concentric whorls of sepals, petals, stamens and carpels. Gary Ditta *et al.* now show that without the concerted encouragement of four such proteins, all that is produced are clusters of leaves (*Curr. Biol.* **14**, 1935–1940; 2004).

The model plant *Arabidopsis thaliana* has more than a hundred genes for MADS-box proteins. Mutations in some of them produce dramatic effects whereas others have redundant functions requiring double,

triple or even more mutants to uncover their roles. The SEPALLATA (SEP) family fall into this latter category, and earned their name when triple mutants — in which three members of the family were all disrupted — produced flowers made entirely of sepals. With this new work, a fourth sibling, SEP4, emerges.

Ditta *et al.* found that mutating SEP4 alone had no obvious effect, but quadruple mutants lacking all four SEP proteins formed 'flowers' with no recognizable floral organs (as shown here). Closer inspection revealed epidermal cells with the same shape and arrangement as in normal leaves. The presence of branched, hair-like structures called trichomes, a feature of *Arabidopsis* leaves, put this



identification beyond doubt.

In his treatise of 1790, *Die Metamorphose der Pflanzen* (*The Metamorphosis of Plants*), the poet and philosopher Johann Wolfgang von Goethe proposed that flowers were modified leaves. Two hundred years later, stripping away the influence of the SEP proteins has returned them to their foliar state.

Christopher Surridge

5. Pardi, L. *Physiol. Zool.* **21**, 1–13 (1948).
6. Rohwer, S. *Evolution* **29**, 593–610 (1975).
7. Senar, J. C. in *Proc. 22 Int. Ornithol. Congr.* (eds Adams, N. J. & Slotow, R. H.) 1669–1686 (1999).
8. Dale, J., Lank, D. B. & Reeve, H. K. *Am. Nat.* **158**, 75–86 (2001).

9. Queller, D. C. *et al. Nature* **405**, 784–787 (2000).
10. West-Eberhard, M. J. *Proc. Am. Phil. Soc.* **123**, 222–234 (1979).
11. Paz-y-Miño C, G., Bond, A. B., Kamil, A. C. & Balda, R. P. *Nature* **430**, 778–781 (2004).
12. Turner, P. E. & Chao, L. *Nature* **398**, 441–443 (1999).

Condensed-matter physics

A magnetic pendulum

Claude Chappert and Thibaut Devolder

Where two oppositely magnetized regions meet, there is a so-called domain wall. Under the right conditions, this wall can be made to oscillate like a pendulum, suggesting a new approach to electronics.

On page 203 of this issue, Saitoh *et al.*¹ provide a beautiful demonstration of a magnetic domain wall oscillating just as a pendulum would do under gravity. A domain wall separates two uniformly magnetized regions and is predicted, at low excitation energies, to behave like a particle with a finite mass. By trapping a single magnetic domain wall in a strip of ferromagnetic material, Saitoh *et al.* show that they can force the oscillations of the wall inside this potential well, achieving resonant displacements of micrometre length at megahertz frequencies — all with only moderate currents, a key result for possible applications.

The experiment designed by Saitoh *et al.* is very clever. They patterned a narrow semicircular loop of $\text{Ni}_{81}\text{Fe}_{19}$ wire, 70 nm wide and 45 nm deep, onto a silicon substrate. In such a wire of soft ferromagnetic material, dipolar interactions tend to maintain the magnetization parallel to the wire edges. A domain wall will thus separate two areas that are uniformly magnetized antiparallel to each other (Fig. 1a). These ‘head-to-head’ walls have a nanoscale core that is magnetized perpendicular to the local wire direction. And this is where the semicircular shape comes into play: a magnetic field strong enough to saturate the magnetization of the whole loop, when reduced again to zero, will either leave a single wall in the loop or destroy a pre-existing one, depending solely on the field’s orientation (Fig. 1b,c). If a wall is created, increasing the field back to moderate values will tend to trap the wall at the position that keeps its core magnetized parallel to the field direction. The wall will then behave in a manner similar to a simple pendulum, which relaxes towards a stable position under gravity. But here the role of gravity is taken by the magnetic field and hence can be tuned in amplitude and direction — gravity itself is not so easy to manipulate!

Applying a periodic force to this ‘magnetic pendulum’ results in an oscillation, whose resonant frequency depends on the effective ‘mass’ of the domain wall: Saitoh *et al.* have determined that mass to be 6.6×10^{-23} kg.

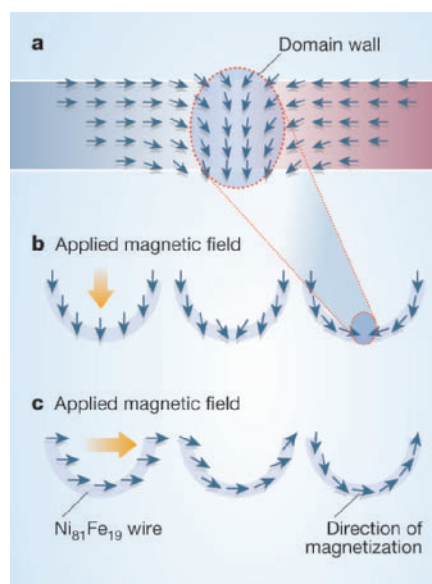


Figure 1 A magnetic pendulum in tunable gravity. a, Where two oppositely magnetized regions meet, there is a domain wall. The arrows indicate the direction of magnetization. Saitoh *et al.*¹ have explored the magnetization in a semicircular loop of $\text{Ni}_{81}\text{Fe}_{19}$ wire. b, When a magnetic field is applied as shown and then removed, the magnetization of the wire settles into a configuration that includes a single domain wall. c, A magnetic field applied in a perpendicular orientation, however, sets the magnetization such that there is no domain wall. When a domain wall is created, Saitoh *et al.* have shown that it can be made to behave like a pendulum does under gravity, although in this case the ‘gravitational’ force — the strength and direction of the magnetic field — can be tuned across a range of values.

The authors have gone further, to investigate the oscillation triggered by passing a sinusoidal current through the wire, using its resonance to explore the interaction between the electrons and the wall. Two types of competing interactions are expected. One is ‘momentum transfer’, which is basically the mechanical pressure exerted by electrons reflected from the domain wall. The other is ‘spin transfer’, which arises from the fact that

the spins of electrons passing through the domain wall rotate: because total angular momentum is conserved, angular momentum is transferred from the electrons to the magnetization itself, which can trigger movement of the domain wall.

To identify which mechanism is the more relevant, Saitoh *et al.* take advantage of the different frequency-dependent efficiencies of each transfer process. They find that momentum transfer is a hundred times more efficient than spin transfer in the frequency range of this experiment (from a few to 75 megahertz). This is unexpected in ferromagnetic metals, such as the nickel–iron compound used here: most scientists believed that domain walls in such materials are wide enough for flowing spins to rotate adiabatically and track the local magnetization orientation, leading to the dominance of spin transfer. Moreover, by properly choosing the excitation frequency, the threshold current to displace the domain wall could be reduced to a few times 10^9 A m^{-2} , a hundredth of what is observed with direct currents, and well below any heating threshold.

Magnetic domain walls can be considered to be self-assembled stable nanostructures; moreover, they can be created or annihilated by external action. The manipulation of domain walls in stripes has already been proposed as a way of storing information or even performing logic functions². With their result, Saitoh *et al.*¹ now provide a low-energy handle on the manipulation of domain walls in complex circuit architectures. And this technique can be fast³ and is fairly scalable — the only requirement is to maintain a constant current density as the stripe dimensions are reduced. Saitoh *et al.* propose some rules of thumb to reduce the required current density even further. In fact, if the ferromagnetic metal were replaced by a dilute magnetic semiconductor, the current density needed to move a wall would already be much smaller⁴.

Conventional (CMOS) electronics is facing major technical hurdles, particularly that of miniaturization. New approaches are needed to reduce energy dissipation and fabrication costs while maintaining the rate of reduction of dimensions, and must be implemented within 15–20 years. This work¹ could well find application in this direction. Saitoh and colleagues could be a step closer to a new ‘domain-wall’ electronics.

Claude Chappert and Thibaut Devolder are at the Institut d’Electronique Fondamentale, UMR CNRS 8622, Université Paris-Sud, 91405 Orsay cedex, France.
e-mail: claudc.chappert@ief.u-psud.fr

1. Saitoh, E., Miyajima, H., Yamaoka, T. & Tataru, G. *Nature* **432**, 203–206 (2004).
2. Allwood, D. *et al. Science* **296**, 2003–2006 (2002).
3. Nakatani, Y., Thiaville, A. & Miltat, J. *Nature Mater.* **2**, 521–523 (2003).
4. Yamanouchi, M., Chiba, D., Matsukura, F. & Ohno, H. *Nature* **428**, 539–542 (2004).

Photochemistry

Selective chemistry in a capsule

J. Am. Chem. Soc. doi:10.1021/ja0450197 (2004)

Chemists have constructed a bowl-shaped molecule that, when it dimerizes in water, traps smaller guest molecules inside and activates them so that they react with light. This 'cavitand' therefore seems to mimic the behaviour of enzymes. Although the cavitand is soluble in water, its inside is dry and hydrophobic, and so attracts certain carbon-based molecules that would otherwise shy away from water.

Lakshmi S. Kaanumalle *et al.* report that when the capsule encloses the small molecules, it causes a very specific light-driven reaction to occur. The cage effect inhibits all but the desired reaction pathway, a selectivity that has previously been seen only in solid-state reactions.

In one particular experiment, the authors note that the cavitand promotes an unusual reaction pathway in which a highly reactive intermediate, called a radical pair, must exist for much longer than normal. They also claim that, during the reaction, part of the guest molecule twists, rearranging its carbon skeleton in a precise way that is controlled by its host.

Mark Peplow

Medicine

Simple SARS detection

Emerg. Infect. Dis. Nov. (2004)

www.cdc.gov/ncidod/EID/vol10no11/04-0516.htm

The outbreak of severe acute respiratory syndrome (SARS) in 2003 left hundreds of people dead. It also left scientists scrambling to find the best way of detecting the virus early on to prevent its spread. The availability of the complete genomic sequence of the coronavirus that causes SARS has allowed rapid detection. But these tests require familiarity with the high-tech equipment for carrying out assays such as reverse transcription and the polymerase chain reaction.

Xiao-Yan Che *et al.* have looked at the advantages of using the nucleocapsid protein, or 'N protein', of the coronavirus as a diagnostic marker. They analysed 420 serum samples taken from 317 SARS patients, and found that the protein could be detected as early as day 1 after the onset of symptoms and until day 18. Using a specially developed immunosorbent assay, they were able to establish N-protein concentrations in the samples. The approach resulted in a test specificity of 99.9%.

The technique could be used in medical centres lacking sophisticated equipment, the authors say, and in mass screening to track the origins of SARS.

Roxanne Khamisi

Microbiology

Sticky insects

Cell 119, 329–341 (2004)

The parasite *Leishmania* causes the sometimes fatal tropical disease leishmaniasis, and is transmitted to humans by the bite of the sand fly. Shaden Kamhawi *et al.* have now identified a protein that helps the pathogen attach to, and survive in, the sand fly's digestive tract.

The researchers screened a library of genes that are active in the midgut of the sand fly *Phlebotomus papatasi*, and found that one gene encodes a member of a family of proteins called galectins, which are involved in insect immunity against pathogens.

Kamhawi *et al.* present evidence that the parasite binds to this protein, called PpGalec, and uses it to latch on to the insect's gut. When the authors fed sand flies with an antibody that binds and blocks PpGalec, the number of parasites in the gut dropped dramatically and the infection was often eliminated altogether.

The authors speculate that using molecules such as PpGalec in vaccines — enabling antibodies against these molecules to be produced in humans or other animal reservoirs — could stop the parasite from gaining a foothold in sand flies after they bite an infected host, and thereby interrupt the transmission cycle.

Helen Pearson

Natural-product chemistry

A toxic trail

Proc. Natl Acad. Sci. USA 45, 15857–15860 (2004)

Native Colombian Indians dip their blow-darts in a poison obtained from the skins of certain species of frog. The major components of the poison, called batrachotoxins, are 250 times more toxic than strychnine. They are not made by the frogs themselves, however, and their apparent dietary origin had remained a mystery.

The same toxins have also been identified in the feathers and skin of birds of the genera *Pitohui* (pictured) and *Ifrita* endemic to New Guinea, where they seem to act as a deterrent to parasites and predators, including humans. Now John P. Dumbacher *et al.* have found out how the birds acquire their batrachotoxins. Following up a suggestion by local people, they identified these toxins in 'nanisani' beetles of the genus *Choresine*. The local word 'nanisani' refers to the tingling and numbing sensation that comes from contact with either these *Choresine* beetles or the *Ifrita* birds; the villagers warned how beetles landing in an eye or on one's face can cause painful burning.



The beetle can't make the complex steroidal carbon skeletons of the batrachotoxins. Presumably, either the beetle, or perhaps some symbiotic organism, modifies steroids acquired from plant sources to make these deadly substances.

Philip Ball

Addiction

Nicotine sensitivity

Science 306, 1029–1032 (2004)

Nerve cells in various parts of the central nervous system bear a receptor for the neurotransmitter acetylcholine. This receptor is also what enables nerves to respond to nicotine — and it is nicotine's addictive properties that are thought to make giving up smoking so difficult.

Twelve different subunits of the receptor have been identified, and they can combine in different ways to make up the pentameric protein. Andrew R. Tapper *et al.* genetically engineered mice to bear a

mutation in the $\alpha 4$ subunit that made the $\alpha 4$ -containing receptors hypersensitive to nicotine. They found that the animals required far lower doses of nicotine than controls to display such addictive symptoms as sensitization, tolerance and a preference for nicotine-containing solutions over saline. It seems that this subunit alone is sufficient to engender addiction.

These mice should, the authors say, be good subjects for further molecular and behavioural studies. The findings also raise the possibility that individual variations in the $\alpha 4$ subunit might underlie people's differing susceptibility to nicotine addiction.

Amanda Tromans

J. P. DUMBACHER

Evolution

How do characters evolve?

Arising from: R. E. Ricklefs *Nature* **430**, 338–341 (2004)

Ricklefs¹ claims to show that morphological evolution in birds is associated with speciation events — that is, it is punctuational — by inference from data on only species number, clade age and character variance from a range of passerine clades. He suggests that variance increases in proportion with clade age under gradual change, but in proportion to the logarithm of species number if change is punctuational. Here I show that both clade age and the logarithm of species number independently predict variance under both gradual and punctuational change, rendering Ricklefs' results uninformative about his central hypothesis.

First, I simulated 100 clades that survived to the same age (60 units) under a constant birth–death process (speciation rate, 0.2; extinction rate, 0.16). Trees contained from one species (excluded from analyses) to 238 species (mean, 49.3). I evolved two characters on each tree under brownian motion — one gradual, the other changing only at speciation. $\log(\text{species number})$ is a highly significant predictor of variance in the evolved trait under both gradual ($t_{96} = 5.42$, $P < 0.0001$, $r^2 = 0.23$) and punctuational change ($t_{96} = 7.57$, $P < 0.0001$, $r^2 = 0.37$), even among clades of the same age.

Second, I evolved 100 surviving clades until they reached a fixed size (50 extant species) by using the same process and evolved characters as before. Clade age ranged from 23.7 to 176.2 time units and significantly (though weakly) predicted variance in the evolved trait under both gradual ($t_{98} = 2.03$, $P = 0.046$, $r^2 = 0.03$) and punctuational ($t_{98} = 2.32$, $P = 0.02$, $r^2 = 0.04$) models, even among clades of the same size.

These results are not surprising. Under gradual change, variance accumulates along phylogenetic branches². Larger clades have more total branch length within them, even in same-aged clades (across my trees, Spearman's $r = 0.98$), and so have more variance in gradually evolved traits. With punctuational change, variance accumulates only at speciations, but older clades have experienced more speciation (and extinction) events — even in same-sized clades — to an extent that depends on the extinction rate³. Clade age is therefore a good predictor of number of speciations across the same-sized trees (Spearman's $r = 0.81$).

Ricklefs¹ found that $\log(\text{species number})$ but not clade age independently predicted morphological variance in multiple regressions, and concludes from this that morphological evolution in birds seems to be associated with cladogenesis. However, my simulations show that his significant results for $\log(\text{species number})$ are expected under

both models. Further, both models predict that clade age too will correlate with variance, but Ricklefs found no significant association. Why not? Ricklefs proposes a model in which speciation promotes gradual divergent change. It is unclear what predictions his model makes about his data. Alternatively, the negative result may simply be a type II error. Clade age was only weakly predictive of variance in my simulations, and the phylogeny from which the clade ages are taken⁴ conflicts markedly with more recent evidence⁵.

The mode of character evolution can sometimes be inferred using detailed phylogenetic information^{3,6,7}. Clade ages and species numbers alone are not enough.

Andy Purvis

Department of Biological Sciences,
Imperial College London, Silwood Park Campus,
Ascot, Berks SL5 7PY, UK
e-mail: a.purvis@imperial.ac.uk
doi:10.1038/nature03092

1. Ricklefs, R. E. *Nature* **430**, 338–341 (2004).
2. Felsenstein, J. *Am. Nat.* **125**, 1–15 (1985).
3. Bokma, F. J. *Evol. Biol.* **15**, 1048–1056 (2002).
4. Sibley, C. G. & Ahlquist, J. E. *Phylogeny and Classification of Birds: A Case Study of Molecular Evolution* (Yale Univ. Press, New Haven, Connecticut, 1990).
5. Barker, F. K., Cibois, A., Schikler, P., Feinstein, J. & Cracraft, J. *Proc. Natl Acad. Sci. USA* **101**, 11040–11045 (2004).
6. Mooers, A. Ø., Vamori, S. M. & Schluter, D. *Am. Nat.* **154**, 249–259 (1999).
7. Pagel, M. *Nature* **401**, 877–884 (1999).

Ricklefs *replies* — Purvis¹ states that, under random gradual change, clades accumulate variance in proportion to their total branch length. Accordingly, at a given age, clades with more species should exhibit greater variance. This is incorrect, as shown both analytically^{2,3} and by simulation^{4,5}, and the error underlines a misunderstanding of my analysis.

Variance depends strictly on the average time of divergence between species within a clade and not on the number of species or total branch length. The positive relationship between species number and variance in Purvis's simulations with fixed time reflects the occurrence of earlier branch points in what turn out to be larger clades. Had Purvis simulated gradual evolution in an unconstrained speciation–extinction process with varied clade ages, he would have obtained a significant partial correlation between variance and time, which I was unable to detect in clades of passerine birds. The weaker time effect in Purvis's second simulation with constrained clade size reflects the distribution of most nodes in each tree at similar depth, regardless of the age of the root.

To determine whether time itself, and hence gradual evolution, contributes to trait

variance, one must analyse data, whether observed or simulated, in which time and the number of species are allowed to vary independently. Purvis's simulations lack this independence: the first does not vary clade age; the second holds the species number constant. The results are not surprising, and they do not address the validity of my analysis. Although average divergence time and number of species are correlated among clades, multiple regression allows one to identify unique contributions of time and species number to trait variance among clades of different ages.

Purvis's comment about a type II error (failing to detect a true relationship) is of more concern, particularly because relative age, based on DNA-hybridization phylogeny⁶, is estimated less well than the number of species. In a test of the monophyly of 40 out of 106 (from ref. 6) tribe-to-family-level clades using a maximum-likelihood analysis of over 4 kilobases of the *RAG-1* and *RAG-2* genes⁷, only three were found that were significantly paraphyletic compared with 27 strongly supported and 10 ambiguous clades⁸. This does not “conflict markedly” with the monophyly of clades used in my analysis. It is of more importance, however, that the relative ages of clades in the sequence-based and DNA-hybridization phylogenies were not compared.

If the phylogeny in ref. 6 provides a reasonably accurate view of clade age, then the absence of a significant time effect on variance among different-aged clades would be sufficient to reject a model of gradual evolution that is independent of species number. Gradual evolutionary divergence, whether fast or slow, driven by interactions among species in a clade (as opposed to punctuated evolution associated with speciation⁹) is also species-dependent, rather than time-dependent, inasmuch as the pressure to diversify is in some way proportional to species number. Increasing knowledge of phylogenetic relationships makes this an opportune time to examine more closely the generation of trait variance in diversifying clades.

Robert E. Ricklefs

Department of Biology, University of Missouri
at St Louis, St Louis, Missouri 63121–4499, USA
e-mail: ricklefs@umsl.edu
doi:10.1038/nature03093

1. Purvis, A. *Nature* doi:10.1038/nature03092 (2004).
2. Slatkin, M. *Paleobiology* **7**, 421–425 (1981).
3. Gavrilits, S. *Proc. R. Soc. Lond. B* **266**, 817–824 (1999).
4. Raup, D. M. & Gould, S. J. *Syst. Zool.* **23**, 305–322 (1974).
5. Valentine, J. W., Collins, A. G. & Meyer, C. P. *Paleobiology* **20**, 131–142 (1994).
6. Sibley, C. G. & Ahlquist, J. E. *Phylogeny and Classification of the Birds of the World* (Yale Univ. Press, New Haven, Connecticut, 1990).
7. Sibley, C. G. & Monroe, B. L. Jr *Distribution and Taxonomy of Birds of the World* (Yale Univ. Press, New Haven, Connecticut, 1990).
8. Barker, F. K., Cibois, A., Schikler, P., Feinstein, J. & Cracraft, J. *Proc. Natl Acad. Sci. USA* **101**, 11040–11045 (2004).
9. Gould, S. J. & Eldredge, N. *Paleobiology* **3**, 115–151 (1977).

Knotted threads of darkness

Dark lines within a laser beam can be manipulated to form stable vortex knots.

Destructive interference may lead to complete cancellation when light waves travelling in different directions cross, and in three-dimensional space this occurs along lines that are vortices of electromagnetic energy flow¹. Here we confirm theoretical predictions^{2,3} by experimentally creating combinations of optical laser beams in which these dark threads form stable loops that are linked and knotted.

Since Kelvin proposed his vortex atom hypothesis⁴, temporally persistent vortex knots have been sought in fluid mechanics^{5,6}, superfluid flow⁷, field theory⁸ and nonlinear excitable media⁹. The optical vortices that we synthesize here are embedded in light beams created by a single-frequency laser (Fig. 1a; for methods, see supplementary information). A direct consequence of the linearity of Maxwell's equations for the electromagnetic field is that the intensity pattern in free space is static, so our optical vortices are stable in both space and time.

The optical phase cannot be defined on these dark threads: it is singular there (like the angle of longitude at the North Pole, for example). Consequently, there is a circulation of electromagnetic energy around the optical vortex. Around such a singularity, the phase changes by an integer multiple of 2π ; this integer, m , is the strength of the singularity.

We combine Laguerre–gaussian beams¹⁰, which have an optical vortex of integral strength m along the beam axis and two further variable parameters: p , the discrete radial number (the number of nodal rings about the beam axis), and w , the continuous waist width (which sets the transverse scale of the beam). These additional degrees of freedom allow us to create the same basic structures as those proposed in refs 2 and 3, but with the vortices more clearly defined by the surrounding intensity.

The loops, links and knot structures arise from the coaxial superposition of four beams. Three of these beams have an on-axis vortex with the same strength m , but different values of p and w . The amplitudes are chosen so that both the field and its radial derivative are zero at a specific radius in the focal plane. The resulting destructive interference forms a dark loop around the axial strength singularity. Adding a simple gaussian beam with small amplitude distorts this loop into an $(m, 2)$ torus knot¹¹, threaded by an m -stranded helix of strength-one vortices. If $m=2$, the vortices form two linked loops (Fig. 1b); if $m=3$, they form a trefoil knot (Fig. 1c).

Our optical beams are produced by using a single-frequency helium–neon laser beam illuminating a hologram. This forms the

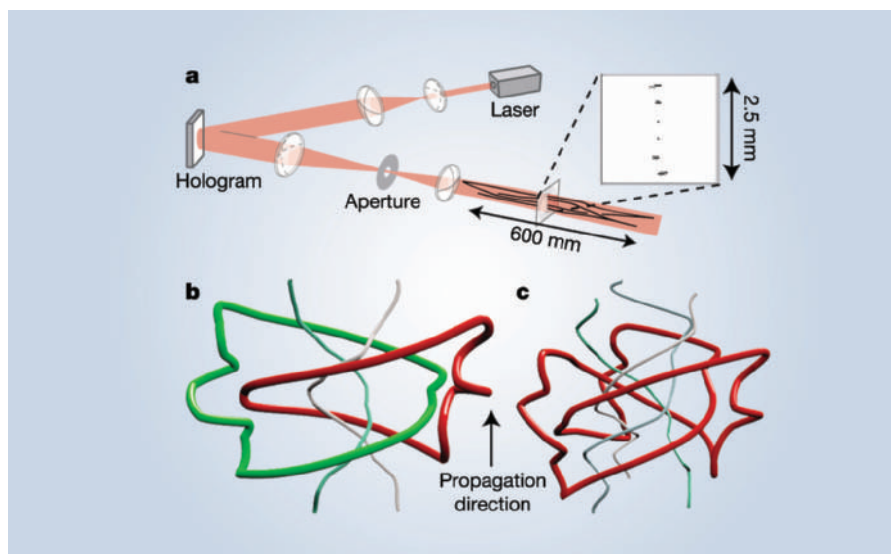


Figure 1 Knotted lines of darkness. **a**, Light from a hologram is passed through a spatial filter and a camera images the beam cross-section in various planes. Thin black lines represent the threads of darkness that form the link or knot. Inset shows the observed over-saturated beam cross-sections at the beam waist of the link; dark points indicate positions where the optical vortex threads cross the observation plane. **b,c**, Three-dimensional representations of measured link (**b**) and trefoil knot (**c**) configurations. The link and knot are threaded by further vortices (represented by the thinner tubes) that follow the axis of the light beam.

intensity and phase structure of the desired superposition of Laguerre–gaussian beams. As the fine structure of the singularities lies in regions of near darkness, it is necessary to oversaturate the camera to determine their positions in the beam cross-section accurately. The resulting image contains points of darkness corresponding to the positions of the dark threads as they intersect the plane of the camera.

We have produced stable link and knot structures formed from vortices in an optical beam. Our observations demonstrate the precision control of light beams that is required to create complex regions of zero light intensity. Such regions could be used, for example, to confine cold atoms and Bose–Einstein condensates in complex topologies¹².

Jonathan Leach*, **Mark R. Dennis†‡**, **Johannes Courtial***, **Miles J. Padgett***

**Department of Physics and Astronomy, University of Glasgow, Glasgow G12 8QQ, UK
e-mail: j.leach@physics.gla.ac.uk*

†H. H. Wills Physics Laboratory, Tyndall Avenue, Bristol BS8 1TL, UK

‡School of Mathematics, University of Southampton, Highfield SO17 1BJ, UK

1. Nye, J. F. & Berry, M. V. *Proc. R. Soc. Lond. A* **336**, 165–190 (1974).
2. Berry, M. V. & Dennis, M. R. *Proc. R. Soc. Lond. A* **457**, 2251–2263 (2001).
3. Berry, M. V. & Dennis, M. R. *J. Phys. A* **34**, 8877–8888 (2001).
4. Thompson, W. *Phil. Mag.* **34**, 15–24 (1867).
5. Moffatt, H. K. *J. Fluid Mech.* **35**, 117–129 (1969).
6. Ricca, R. L., Samuels, D. C. & Barenghi, C. F. *J. Fluid Mech.* **391**, 29–44 (1999).
7. Poole, D. R., Scofield, H., Barenghi, C. F. & Samuels D. C. *J. Low Temp. Phys.* **132**, 97–117 (2003).
8. Fadeev, L. & Niemi, A. J. *Nature* **387**, 58–61 (1997).
9. Winfree, A. T. *Nature* **371**, 233–236 (1994).
10. Allen, L., Beijersbergen, M. W., Spreeuw, R. J. C. & Woerdman, J. P. *Phys. Rev. A* **45**, 8185–8189 (1992).
11. Adams, C. C. *The Knot Book* (Freeman, New York, 1994).
12. Ruostekoski, J. & Anglin, J. R. *Phys. Rev. Lett.* **86**, 3934–3936 (2001).

- Supplementary information accompanies this communication on Nature's website.
Competing financial interests: declared none.

Plant genetics

Gene transfer from parasitic to host plants

Plant mitochondrial genes are transmitted horizontally across mating barriers with surprising frequency, but the mechanism of transfer is unclear^{1,2}. Here we describe two new cases of horizontal gene transfer, from parasitic flowering plants to their host flowering plants, and present phylogenetic and biogeographic evidence that this occurred as a result of direct physical contact between the two. Our findings complement the discovery that genes can be transferred in the opposite direction, from host to parasite plant³.

Both cases of horizontal gene transfer involve the mitochondrial gene *atp1* and the recipient *Plantago*, a large cosmopolitan genus comprising mostly weeds. We examined 43 *Plantago* species, all containing an intact *atp1* gene. Despite the exceptional divergence of these genes, their relationships

to one another (Fig. 1, top) and their inclusion in Plantaginaceae agree with independent estimates⁴ of the phylogeny for these taxa. This clade of *Plantago atp1* genes is therefore most likely to be the product of standard vertical transmission.

Each of three related *Plantago* species contains a second, full-length *atp1* copy with several frameshift mutations. This clade of pseudogenes is allied with the distantly related genus *Cuscuta*, as indicated by bootstrap (97%) and other analyses (Fig. 1, and see supplementary information), implying that a species of *Cuscuta* transferred a copy of *atp1* to the common ancestor of these three *Plantago* species. *Cuscuta* is a large genus of flowering plants (also known as dodders; Fig. 2) that parasitizes an unusually broad range of hosts⁵, including *Plantago*⁶. The three *Plantago* species are native to Europe and north Africa, as are many dodders, including *C. europaea* (Fig. 1).

Two other species of *Plantago* (*P. rigida* and *P. tubulosa*) also contain a second, full-length copy of *atp1*. These too are pseudogenes and are probably also derived by horizontal gene transfer. *Bartsia*, a genus in the parasitic family Orobanchaceae, is identified as the likely gene donor by phylogenetic analysis (81% bootstrap support; Fig. 1). Biogeography neatly supports this inference, because *P. rigida*, *P. tubulosa* and most species of *Bartsia* are found exclusively at high altitudes in the northern Andes^{7,8}.

Both lineages of *Plantago atp1* pseudogenes were almost certainly acquired by horizontal gene transfer, as indicated by evidence against their inclusion in the *Plantago*-wide clade of vertically transmitted *atp1* genes and in favour of their phylogenetically anomalous inclusion with parasitic plants. (The possibility of DNA contamination can be ruled out; see supplementary information.) Phylogenetic distribution of the transferred genes and molecular clock analysis (see supplementary information) indicate that both transfers were recent (within the past few million years).

Both transfers probably occurred by direct plant-to-plant transmission of DNA from parasitic to host plants. This is indicated by phylogenetic evidence that parasitic plants, which penetrate host plants intracellularly as part of their normal life cycle, served as donors in both transfers. It is also suggested by the striking biogeographic concordance of donors and recipients in the case of the high-altitude, northern Andean *Bartsia*, *P. rigida* and *P. tubulosa*, as well as by the broad host ranges of both groups of parasitic donors. The well documented exchange in both directions of macromolecules, viruses and phytoplasmas between parasitic plants and their hosts^{9–11} is also consistent with plant-to-plant transmission.

We cannot altogether rule out the possibility that the involvement of parasitic plants

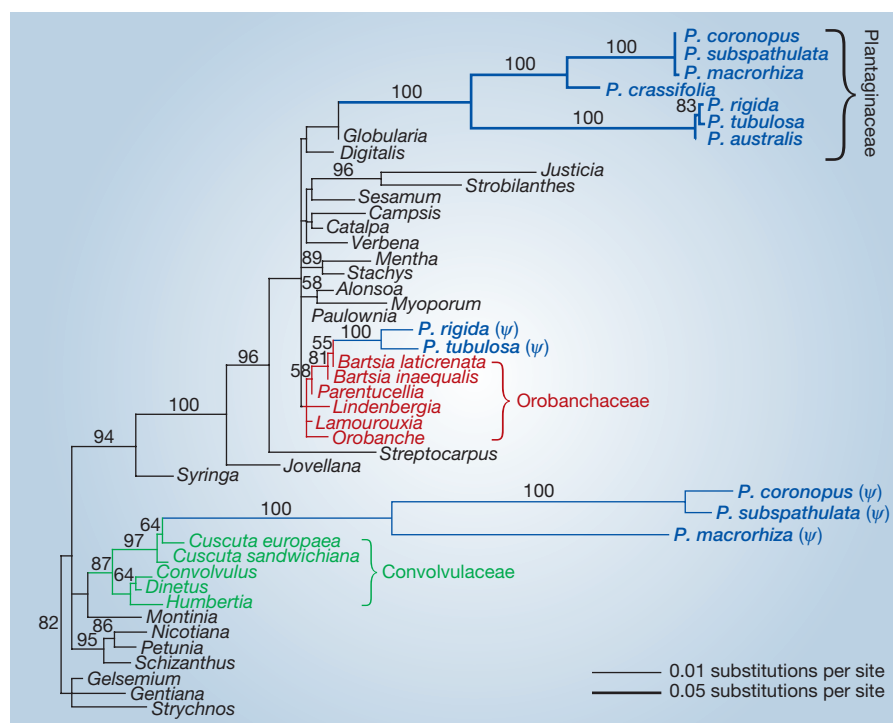


Figure 1 Phylogenetic evidence for two horizontal transfer events of the gene *atp1* into *Plantago* (blue). Seven *Plantago atp1* genes at the top of the maximum-likelihood tree are intact, vertically transmitted, and rapidly evolving (scale reduced by 80%). The other two sets of *Plantago atp1* genes are pseudogenes (ψ) acquired from parasitic plants in the Orobanchaceae (red) and Convolvulaceae (green). Bootstrap values of over 50% are shown. For methods, see supplementary information.

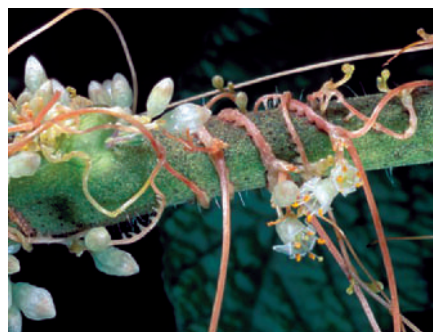


Figure 2 A parasitic dodder (*Cuscuta californica*) in flower, with its haustoria penetrating a host tomato plant.

as donors is fortuitous — for example, some biological vector (such as pollen, fungus, a bacterium or virus, or an insect) may have caused indirect transfer. But our direct-transfer hypothesis is more parsimonious and is supported by the evidence available.

The parasitic lifestyle has evolved roughly a dozen times in flowering plants and there are more than 4,000 species of parasitic plants¹², providing plenty of opportunity for horizontal gene transfer not only from parasite to host, but also from host to parasite³. Parasites with narrow host ranges may exchange genes frequently with their hosts. Those with broad host ranges (dodders, for example) may also act as vectors for gene transfer by directly bridging unrelated host species. In addition to parasitism, other vectors, and non-vectored mechanisms such as grafting, may promote horizontal gene transfer during plant evolution.

Jeffrey P. Mower, Saša Stefanović*, Gregory J. Young, Jeffrey D. Palmer
Department of Biology, Indiana University,
Bloomington, Indiana 47405, USA
e-mail: jpalmer@bio.indiana.edu

*Present address: Department of Biology,
University of Toronto at Mississauga,
Mississauga, Ontario L5L 1C6, Canada

- Bergthorsson, U., Adams, K. L., Thomason, B. & Palmer, J. D. *Nature* **424**, 197–201 (2003).
- Won, H. & Renner, S. S. *Proc. Natl Acad. Sci. USA* **100**, 10824–10829 (2003).
- Davis, C. C. & Wurdack, K. J. *Science* **305**, 676–678 (2004).
- Rønsted, N., Chase, M. W., Albach, D. C. & Bello, M. A. *Bot. J. Linn. Soc.* **139**, 323–338 (2002).
- Yuncker, T. G. *Mem. Torrey Bot. Club* **18**, 111–331 (1932).
- Tessene, M. F. *Mich. Bot.* **8**, 72–104 (1969).
- Molau, U. *Opera Bot.* **102**, 1–99 (1990).
- Luteyn, J. *Parasites: A Checklist of Plant Diversity, Geographical Distribution, and Botanical Literature* (New York Botanical Garden Press, Bronx, New York, 1999).
- Hosford, R. M. *Bot. Rev.* **33**, 387–406 (1967).
- Marcone, C., Hergenrother, E., Ragozzino, A. & Seemüller, E. *J. Phytopathol.* **147**, 187–192 (1999).
- Haupt, S., Oparka, K. J., Sauer, N. & Neumann, S. J. *Exp. Bot.* **52**, 173–177 (2001).
- Nickrent, D. L. et al. in *Molecular Systematics of Plants II* (eds Soltis, D. E., Soltis, P. S. & Doyle, J. J.) 211–241 (Kluwer, Boston, 1998).

Supplementary information accompanies this communication on Nature's website.

Competing financial interests: declared none.

brief communications arising online

www.nature.com/bca

Evolution: How do characters evolve?

A. Purvis (doi:10.1038/nature03092)

Reply: R. E. Ricklefs (doi:10.1038/nature03093)

Evolution

How do characters evolve?

Arising from: R. E. Ricklefs *Nature* **430**, 338–341 (2004)

Ricklefs¹ claims to show that morphological evolution in birds is associated with speciation events — that is, it is punctuational — by inference from data on only species number, clade age and character variance from a range of passerine clades. He suggests that variance increases in proportion with clade age under gradual change, but in proportion to the logarithm of species number if change is punctuational. Here I show that both clade age and the logarithm of species number independently predict variance under both gradual and punctuational change, rendering Ricklefs' results uninformative about his central hypothesis.

First, I simulated 100 clades that survived to the same age (60 units) under a constant birth–death process (speciation rate, 0.2; extinction rate, 0.16). Trees contained from one species (excluded from analyses) to 238 species (mean, 49.3). I evolved two characters on each tree under brownian motion — one gradual, the other changing only at speciation. $\log(\text{species number})$ is a highly significant predictor of variance in the evolved trait under both gradual ($t_{96} = 5.42$, $P < 0.0001$, $r^2 = 0.23$) and punctuational change ($t_{96} = 7.57$, $P < 0.0001$, $r^2 = 0.37$), even among clades of the same age.

Second, I evolved 100 surviving clades until they reached a fixed size (50 extant species) by using the same process and evolved characters as before. Clade age ranged from 23.7 to 176.2 time units and significantly (though weakly) predicted variance in the evolved trait under both gradual ($t_{98} = 2.03$, $P = 0.046$, $r^2 = 0.03$) and punctuational ($t_{98} = 2.32$, $P = 0.02$, $r^2 = 0.04$) models, even among clades of the same size.

These results are not surprising. Under gradual change, variance accumulates along phylogenetic branches². Larger clades have more total branch length within them, even in same-aged clades (across my trees, Spearman's $r = 0.98$), and so have more variance in gradually evolved traits. With punctuational change, variance accumulates only at speciations, but older clades have experienced more speciation (and extinction) events — even in same-sized clades — to an extent that depends on the extinction rate³. Clade age is therefore a good predictor of number of speciations across the same-sized trees (Spearman's $r = 0.81$).

Ricklefs¹ found that $\log(\text{species number})$ but not clade age independently predicted morphological variance in multiple regressions, and concludes from this that morphological evolution in birds seems to be associated with cladogenesis. However, my simulations show that his significant results for $\log(\text{species number})$ are expected under

both models. Further, both models predict that clade age too will correlate with variance, but Ricklefs found no significant association. Why not? Ricklefs proposes a model in which speciation promotes gradual divergent change. It is unclear what predictions his model makes about his data. Alternatively, the negative result may simply be a type II error. Clade age was only weakly predictive of variance in my simulations, and the phylogeny from which the clade ages are taken⁴ conflicts markedly with more recent evidence⁵.

The mode of character evolution can sometimes be inferred using detailed phylogenetic information^{3,6,7}. Clade ages and species numbers alone are not enough.

Andy Purvis

Department of Biological Sciences,
Imperial College London, Silwood Park Campus,
Ascot, Berks SL5 7PY, UK
e-mail: a.purvis@imperial.ac.uk
doi:10.1038/nature03092

1. Ricklefs, R. E. *Nature* **430**, 338–341 (2004).
2. Felsenstein, J. *Am. Nat.* **125**, 1–15 (1985).
3. Bokma, F. J. *Evol. Biol.* **15**, 1048–1056 (2002).
4. Sibley, C. G. & Ahlquist, J. E. *Phylogeny and Classification of Birds: A Case Study of Molecular Evolution* (Yale Univ. Press, New Haven, Connecticut, 1990).
5. Barker, F. K., Cibois, A., Schikler, P., Feinstein, J. & Cracraft, J. *Proc. Natl Acad. Sci. USA* **101**, 11040–11045 (2004).
6. Mooers, A. Ø., Vamori, S. M. & Schluter, D. *Am. Nat.* **154**, 249–259 (1999).
7. Pagel, M. *Nature* **401**, 877–884 (1999).

Ricklefs *replies* — Purvis¹ states that, under random gradual change, clades accumulate variance in proportion to their total branch length. Accordingly, at a given age, clades with more species should exhibit greater variance. This is incorrect, as shown both analytically^{2,3} and by simulation^{4,5}, and the error underlines a misunderstanding of my analysis.

Variance depends strictly on the average time of divergence between species within a clade and not on the number of species or total branch length. The positive relationship between species number and variance in Purvis's simulations with fixed time reflects the occurrence of earlier branch points in what turn out to be larger clades. Had Purvis simulated gradual evolution in an unconstrained speciation–extinction process with varied clade ages, he would have obtained a significant partial correlation between variance and time, which I was unable to detect in clades of passerine birds. The weaker time effect in Purvis's second simulation with constrained clade size reflects the distribution of most nodes in each tree at similar depth, regardless of the age of the root.

To determine whether time itself, and hence gradual evolution, contributes to trait

variance, one must analyse data, whether observed or simulated, in which time and the number of species are allowed to vary independently. Purvis's simulations lack this independence: the first does not vary clade age; the second holds the species number constant. The results are not surprising, and they do not address the validity of my analysis. Although average divergence time and number of species are correlated among clades, multiple regression allows one to identify unique contributions of time and species number to trait variance among clades of different ages.

Purvis's comment about a type II error (failing to detect a true relationship) is of more concern, particularly because relative age, based on DNA-hybridization phylogeny⁶, is estimated less well than the number of species. In a test of the monophyly of 40 out of 106 (from ref. 6) tribe-to-family-level clades using a maximum-likelihood analysis of over 4 kilobases of the *RAG-1* and *RAG-2* genes⁷, only three were found that were significantly paraphyletic compared with 27 strongly supported and 10 ambiguous clades⁸. This does not “conflict markedly” with the monophyly of clades used in my analysis. It is of more importance, however, that the relative ages of clades in the sequence-based and DNA-hybridization phylogenies were not compared.

If the phylogeny in ref. 6 provides a reasonably accurate view of clade age, then the absence of a significant time effect on variance among different-aged clades would be sufficient to reject a model of gradual evolution that is independent of species number. Gradual evolutionary divergence, whether fast or slow, driven by interactions among species in a clade (as opposed to punctuated evolution associated with speciation⁹) is also species-dependent, rather than time-dependent, inasmuch as the pressure to diversify is in some way proportional to species number. Increasing knowledge of phylogenetic relationships makes this an opportune time to examine more closely the generation of trait variance in diversifying clades.

Robert E. Ricklefs

Department of Biology, University of Missouri
at St Louis, St Louis, Missouri 63121–4499, USA
e-mail: ricklefs@umsl.edu
doi:10.1038/nature03093

1. Purvis, A. *Nature* doi:10.1038/nature03092 (2004).
2. Slatkin, M. *Paleobiology* **7**, 421–425 (1981).
3. Gavrilits, S. *Proc. R. Soc. Lond. B* **266**, 817–824 (1999).
4. Raup, D. M. & Gould, S. J. *Syst. Zool.* **23**, 305–322 (1974).
5. Valentine, J. W., Collins, A. G. & Meyer, C. P. *Paleobiology* **20**, 131–142 (1994).
6. Sibley, C. G. & Ahlquist, J. E. *Phylogeny and Classification of the Birds of the World* (Yale Univ. Press, New Haven, Connecticut, 1990).
7. Sibley, C. G. & Monroe, B. L. Jr. *Distribution and Taxonomy of Birds of the World* (Yale Univ. Press, New Haven, Connecticut, 1990).
8. Barker, F. K., Cibois, A., Schikler, P., Feinstein, J. & Cracraft, J. *Proc. Natl Acad. Sci. USA* **101**, 11040–11045 (2004).
9. Gould, S. J. & Eldredge, N. *Paleobiology* **3**, 115–151 (1977).

Mantle segmentation along the Oman ophiolite fossil mid-ocean ridge

Laurent Le Mée, Jacques Girardeau & Christophe Monnier

UMR-CNRS 6112, Laboratoire de Planétologie et Géodynamique, Université de Nantes, 2 rue de la Houssinière, BP 92208, 44322 Nantes cedex 3, France

It has been difficult to relate the segmentation of mid-ocean ridges to processes occurring in the Earth's underlying mantle, as the mantle is rarely sampled directly and chemical variations observed in lavas at the surface are heavily influenced by details of their production as melt extracted from the mantle. Our understanding of such mantle processes has therefore relied on the analysis of pieces of fossil oceanic lithosphere now exposed at the Earth's surface, known as ophiolites. Here we present the phase chemistry and whole-rock major- and trace-element contents of 174 samples of the mantle collected along over 400 km of the Oman Sultanate ophiolite. We show that, when analysed along the fossil ridge, variations of elemental ratios sensitive to the melting process define a three-dimensional geometry of mantle upwellings, which can be related to the segmentation observed in modern mid-ocean ridge environments.

The morphological segmentation of present oceanic ridges^{1–5} allows us to define different segmentation orders as reported in the general ridge segmentation model⁴. The first-order discontinuities (transform faults) limit large-scale upwellings of deep asthenospheric domes that control the mantle partial melting and therefore the magma production at the origin of the oceanic crust. They are themselves divided into second-order segments limited by non-transform discontinuities (NTD) and overlapping spreading centres (OSC), in which smaller but more depleted mantle diapirs rise up to shallow levels. However, this general model, suitable for oceanic domains away from hotspots, does not explain the chemical variations of the basalts from the Mid-Atlantic Ridge just south of the Açores hotspot—the compositions of these basalts would match primary variations in the mantle composition better than partial melting amounts^{6–8}.

Tests of this segmentation model using mantle rocks can effectively be done in the Oman ophiolite fossil oceanic lithosphere, which displays palaeo-ridge remnants over large distances (420 km)⁹, and for which evidence for along-ridge segmentation has already been described on the basis of structural features. But the ridge arguments were based on structural features only, such as (1) changes in the sheeted dyke trend^{9–11}, (2) the presence of dyke penetration zones that are supposed to mark magmatic pulse tips¹⁰, (3) thickness variations of the crustal and transition zone units^{12,13}, and (4) the presence of mantle diapiric structures defined on the basis of peridotite petrofabrics^{9,14–16}. These criteria make it possible to define four segments, 50 to 100 km long, separated by second-order discontinuities, including at least two OSCs^{9,10}. However, this was not supported by any chemical data, neither on the lavas as in present oceanic ridges, nor directly on the mantle component (for which characteristics are given in this paper).

Sample origin

The Oman ophiolite (Fig. 1a and b) exhibits an exceptionally well-preserved palaeo-oceanic lithosphere, exposed in poorly dislocated massifs that have been put together to reconstruct a single lithosphere segment about 420 km long by 40 km wide⁹. It comprises a crustal unit, 2 to 5 km thick, that dominantly outcrops in the eastern part of the massifs. In this ophiolite, the axis of a northwest–southeast palaeo-ridge has been located using the dyke swarm trend and the presence of frozen structures thought to have been acquired beneath an oceanic ridge⁹. The recognition of dyke penetration zones and of magma chamber terminations^{10,17} and of most of the mantle diapirs^{13–15} suggested that the ridge itself was obducted.

Because our samples were all collected at a distance less than 10 km from the inferred palaeo-ridge, we considered the possibility that, to a first approximation, they may have been accreted over the same span of time. To get around this hypothesis, we also tested the across-strike chemical variability of our sampling, which is extremely low, as shown in the 'Chemistry results' section.

The mantle unit, 5 to 10 km in thickness, dominantly outcrops in the western part of the reconstructed ophiolite, that is, in zones where the palaeo-Mohorovičić discontinuity (Moho) dips in general 50 degrees to the east. It consists mainly of homogeneous harzburgites, except near the crust–mantle transition zone, where it displays more dunitic and wehrlitic compositions, and near the basal thrust contact, where it is locally enriched in clinopyroxene¹⁸. The peridotites usually display decimetre-thick orthopyroxenitic and dunitic bandings parallel to the main foliation. The peridotites are also cut by dykes of different compositions that have been classified into two main suites. The first dykes, covering about 25% of the mantle section, are derived from mid-ocean-ridge basalt (MORB)-like melts. They are interpreted as frozen indicators of diapiric magmatism activity¹⁷. The second dykes are silica-rich and water-rich melts subsequent to the MORB-like suite, "coming from a shallow and partly hydrated lithosphere residual after MORB"¹⁷. To avoid the presence of contamination linked to late impregnation and/or metamorphic processes, we selected samples that lay away from bandings and dykes that could be seen with the naked eye. The samples were also taken at a vertical distance of up to 500 m from the basal thrust and Moho and, as much as possible, in the lesser serpentinized rocks having coarse-grained textures and harzburgitic compositions.

The mantle rocks were sampled during the 2000 and 2001 field campaigns. This amounts to about 280 samples, taken with an approximate 1.5-km increment, from the Wadi Tayin massif that marks the zero on the distance scale used in this work (Fig. 1b), to the Khawr-Fakkan massif to the north. In the laboratory, each sampling site was vertically projected on the palaeo-ridge line. Hence, each sample is represented by a point on the longitudinal profile. However, because of difficulties with access in some areas, samples were not all chosen at the same vertical distances to the palaeo-Moho. We therefore verified that our sampling was representative of the whole mantle on a vertical section perpendicular to the palaeo-ridge, which is discussed in the 'Chemistry results' section.

Petrographic data

Petrographic and textural investigations have been done on each

sample using classical optic procedures and numerical image analyses for the determination of the modal content. It has confirmed that all rocks are medium- to coarse-grained porphyroclastic harzburgites with 10 to 30% secondary serpentine minerals replacing olivine porphyroclasts. The peridotites contain orthopyroxenes (opx) that locally form isolated coarse crystals with common resorbed shapes or that form millimetre-thick aggregates. We have shown that minerals from these aggregates are in thermodynamic

equilibrium with other primary phases, and consider that they derived from former thin opx aggregates, orthopyroxenitic layers or dykelets, tectonically dispersed during their plastic deformation at the origin of the peridotite porphyroclastic structure. This process has increased the opx amount observed in about 20% of the samples, so that the opx value is 15%. About 30% of the harzburgites contains minor (<3%) clinopyroxene (cpx), whereas others only display traces of cpx, generally less than one millimetre in size

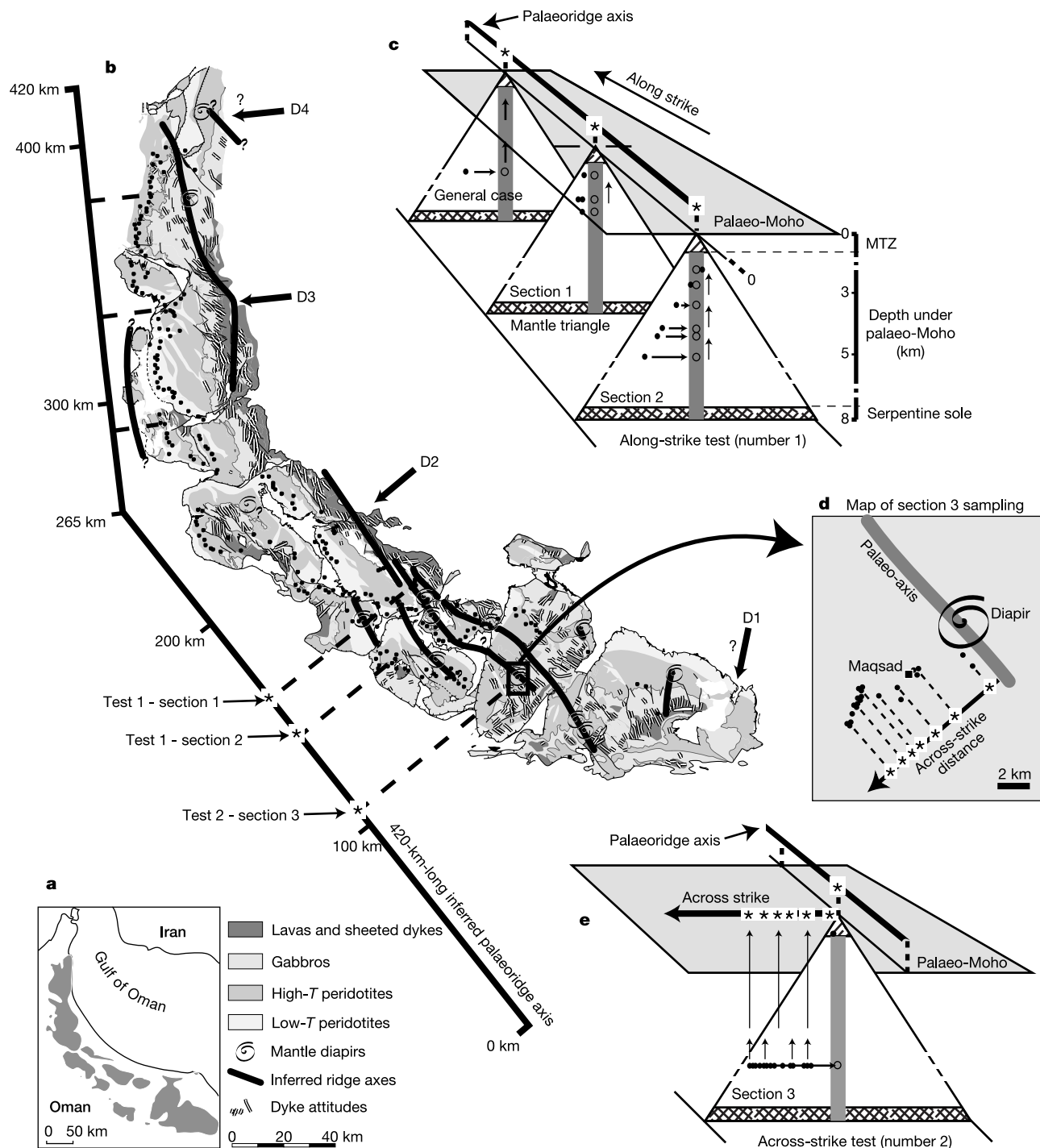


Figure 1 Sampling location and projection method. **a, b**, Location of the 280 samples taken within the Oman mantle (map after ref. 9). Each sample defines a specific mantle triangle orthogonal to the inferred 420-km-long palaeo-ridge. D1 to D4 indicate the four major segments defined. **c**, Projection method ('General case'): each sample (full circle) has been vertically projected on the ridge axis plane (open circle), so that it is represented

by only one point on the ridge (star). Sections 1 and 2 illustrate cases where, in a single triangle, several samples were collected at increasing distances from the palaeo-Moho to test the vertical homogeneity of the mantle. **d**, Location of the across-strike section-3 samples in the Maqсад area. **e**, Projection method for the across-strike test.

and located at the periphery of the opx crystals. Spinel (3% maximum) commonly presents a holly-leaf shape and is located at the periphery or within the opx porphyroclasts, which is known to be primary and acquired during partial melting^{19,20}. More rarely, spinels also constitute subeuhedral crystals at olivine grain boundaries.

Chemistry results

To describe the chemistry of the Oman harzburgites, more than 3,000 analyses were carried out at the cores of their primary phases using a SX50 CAMEBAX microprobe in 174 representative samples. Among these, we obtained 90 bulk-rock major analyses using the ICP-AES laboratory in Brest and 158 trace element analyses from

the ICP-MS laboratory in Toulouse (Supplementary Table 1).

Our data on the primary phases, and particularly the Fe^{2+}/Mg and Cr/Al ratios between holly-leaf-shaped spinel, pyroxenes and olivine, indicate that these phases are in equilibrium. They also show that they follow normal partial melting trends as already seen in most oceanic and ophiolitic peridotites^{21,22}. Most of the peridotites are strongly refractory rocks, as shown (1) by the high values of the $\text{Cr}/(\text{Cr} + \text{Al})$ ratio of primary spinels (down to 0.2, but between 0.45 and 0.67 for 71% of the samples; Fig. 2a), (2) by their MgO bulk-rock content (average 45%), (3) by their very low rare-earth elements (REE) concentrations (normalized Yb up to 0.77, but between 0.08 and 0.28 for 79% of the samples; Fig. 2b) and (4) by

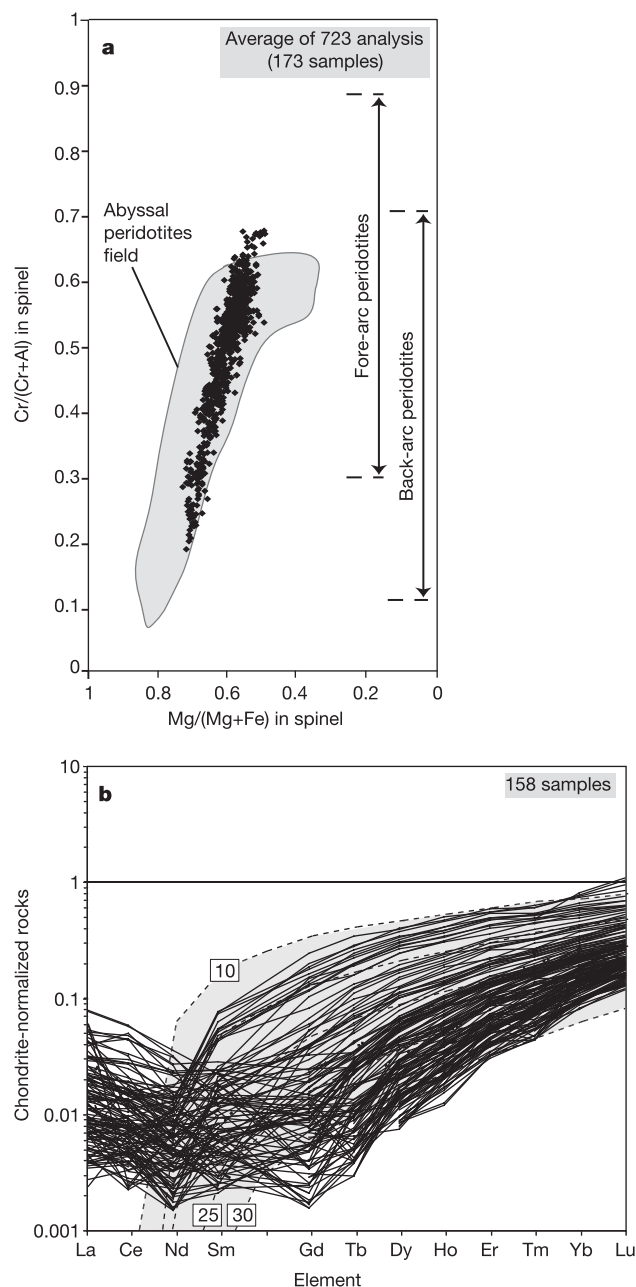


Figure 2 Mineral and whole-rock chemical variation range. **a**, $\text{Cr}/(\text{Cr} + \text{Al})$ versus $\text{Mg}/(\text{Mg} + \text{Fe})$ in spinels. Fields concerning abyssal, fore-arc and back-arc peridotites are from different compilations (refs 47, 48 and 49 respectively). **b**, Chondrite-normalized REE patterns for the study peridotites (chondrite concentrations from ref. 50). The dotted lines are patterns for 10 to 30% melt extraction (after ref. 25).

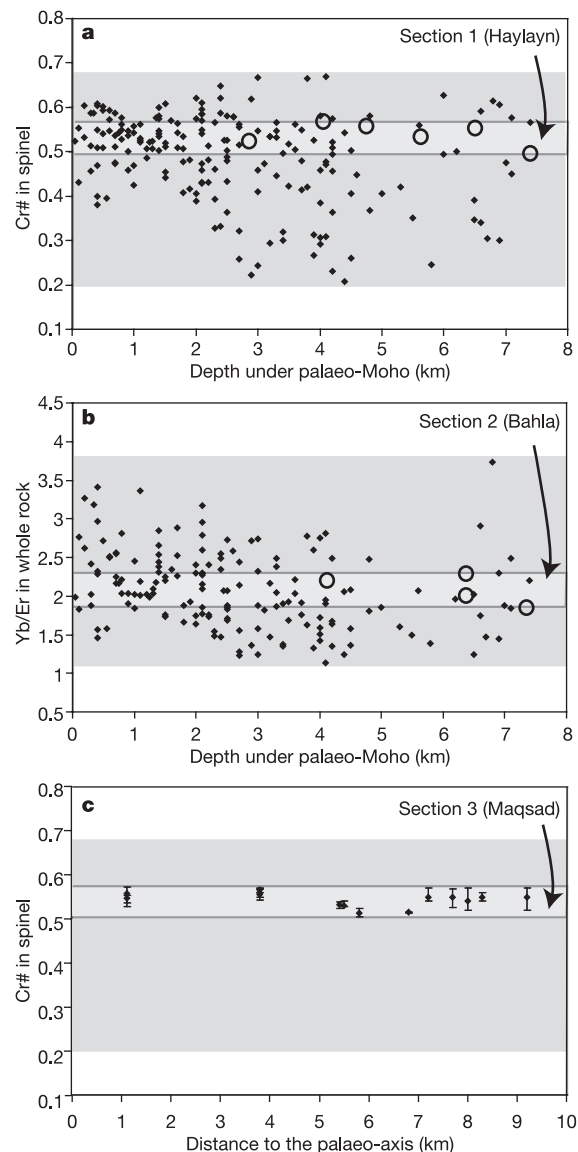


Figure 3 Depth and across-strike mantle homogeneity tests. **a**, **b**, $\text{Cr}/(\text{Cr} + \text{Al})$ in spinel (**a**) and Yb/Er ratio in whole rock (**b**) versus depth under palaeo-Moho (in km) for all samples (small dots) and for specific samples (open large circles, beginning at number 175 on the right, see Supplementary Table 1) collected on sections 1 and 2 (see location on Fig. 1b, c). The dark grey areas display the chemical variability with depth for all studied samples and the light grey areas the variability observed in single mantle triangles as exemplified by sections 1 and 2. **c**, $\text{Cr}/(\text{Cr} + \text{Al})$ in spinel versus distance to the palaeoaxis (in km) for samples collected from section 3 (see location on Fig. 1d, first sample on the right: sample number 185 in Supplementary Table 1). The error bars represent the variability at the scale of the sample or at the scale of the sampling site (comprising in general three samples). Data from the Haylayn, Bahla and Maqsad massifs are indicated.

the Yb/Er ratios that range from 1.1 to 3.7, as also observed in the Sumail and Wuqba peridotites^{23,24}. Their refractory character is also illustrated by the shapes of the REE patterns, which are dominantly characterized by strongly decreasing trace-element contents from the HREE (heavy REE) to the MREE (middle REE). Using previous estimations²⁵, the REE concentrations indicate that peridotites underwent between 10 and 30% melt extraction, which is similar to the values calculated from spinel compositions using equations in refs 26 and 27 for fractional melting^{28–31}. Note that the studied peridotites are also slightly enriched in LREE (light REE). But these enrichments display no correlation either with the quantity of serpentine (deduced from the ICP-AES measured loss on ignition) or with the different melting indicators used, which indicates that the rocks underwent some metasomatism^{28,32}. However, as have other studies^{28,29,31}, we assume that this metasomatism has not affected the MREE and HREE if it has, (or to the same extent) which implies that the Yb/Er slopes have been conserved³³.

Our data show that all studied melting parameters display particularly large variations at the ophiolite scale. But we note that these large chemical variabilities are well organized at the ophiolite scale, when examined along the ridge.

Vertical and across-strike axis chemical variability

We investigated the origin of the observed variations: do they relate to the peridotite partial melting history or do they simply correspond to differences existing in the composition of the parental protoliths? That all studied peridotites are roughly comparable in terms of protolith, except for the rocks that have been mechanically enriched in opx, suggests that the observed chemical variations do reflect the melting history, unlike the opx modal content, which we therefore cannot consider to be a melting indicator in this case³⁴.

We wondered whether the melting relates to the initial accretion event at an oceanic ridge or in a subduction zone context. According to the oxygen fugacity, which can reveal the presence of water and which was calculated for the spinel phases using the olivine–spinel oxybarometer of ref. 35, our peridotites suffered relatively low

oxidation, much like the present-day reduced oceanic mantle³⁶. The data do not show any correlation between the oxygen fugacity and Cr# in spinel, in contrast to what is seen after channelling of subduction zone melts in reduced oceanic mantle, for example in the Izu-Bonin-Mariana area^{22,37}. These arguments instead point to anhydrous melting of an oceanic reduced mantle with no obvious melt–rock interactions for the studied peridotites.

To verify whether our sampling was representative of the whole mantle on a vertical section beneath the palaeo-ridge, we have analysed the chemical variations along two profiles (sections 1 and 2, Fig. 1c), along which samples were picked at increasing distances, up to 8 km from the palaeo-Moho. The data (Fig. 3a, b) show that the chemical variation measured from one sample to another is not depth-dependent, because it is very small (5 to 10 times less) compared to what is observed at the ophiolite scale. At a given sampling site, a rock can therefore be considered as representative of the whole mantle along a vertical section—which is not surprising, because it is supposed to represent a relatively narrow zone from which melt was intensively extracted to produce the crust^{4,38}.

The across-strike homogeneity of the mantle was also tested on one section from the Maqсад area (Fig. 1d, e), where the palaeo-Moho is nearly flat and where the mantle was densely sampled. Our data show that the across-strike chemical variations are very low, as seen for instance by the Cr# compositions of primary spinels, for

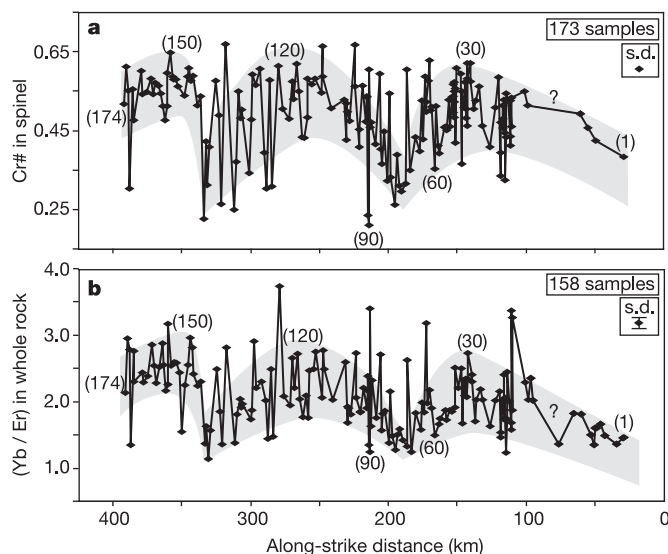


Figure 4 Along-ridge variations of the raw chemistry of the studied peridotites. **a, b**, Variations of the Cr/(Cr + Al) in spinel (**a**) and the Yb/Er ratio in whole rock (**b**) without any statistical processing. 90% of samples are situated in the grey area. s.d., the standard deviation including error analysis and within sample variability. D1 to D4: as for Fig. 5. Six of the samples located between 0 and 75 km on 4b are from ref. 24. Samples are labelled directly on the signal from 1 to 174 (see Supplementary Table 1) from right to left of the figure.

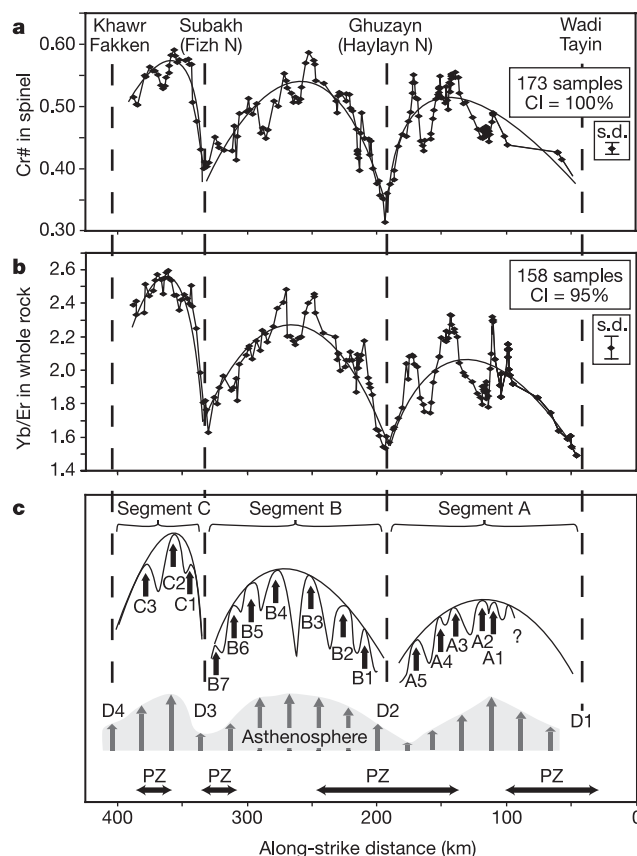


Figure 5 Along-ridge variations of the peridotite chemistry after processing and the resulting synthetic model for the Oman mantle segmentation. **a, b**, Variations of the Cr/(Cr + Al) in spinel (**a**) and the Yb/Er ratio in whole rock (**b**). CI, the correlation index between peaks and valleys on the two figures. **c**, Synthetic model for the Oman mantle segmentation, showing the three 100- to 160-km-long segments (A, B and C), thought to represent large asthenospheric domes, limited by four major discontinuities (D1 to D4). A1, A2... mark the shortest (10–20 km) wavelength variations, interpreted as small mantle diapirs and feeding zones for the magmatic crust. PZ locates the dyke penetration zones defined by ref. 10. s.d., as for Fig. 4.

which the compositional range is nearly six times less than that determined at the ophiolite scale (Fig. 3c). This suggests that a peridotite can be considered to be representative of the mantle accreted at the ridge axis, even if it has been taken from an off-axis site.

Along-axis chemical variability

On Fig. 4, we present the raw variations, without any statistical processing, of two partial melting chemical tracers versus the longitudinal profile on which each studied sample is located (Fig. 1b). Parameters are the Cr/(Cr + Al) ratio in primary spinels, considered to be a very good tracer of partial melting^{21,39,40}, and the bulk-rock Yb/Er ratio, the slope of which is strongly dependent on the melting rate³³. They both display pronounced chemical variations described by peaks and valleys, with a relatively short (5–12 km long) wavelength and variable amplitudes. The lowest values of the melting indexes concentrate on narrow areas, around points 190 and 330 km on the ridge scale (shown in Fig. 1b), which allows us to define three larger zones with higher values and the highest points at the centre of the zones. These zones correspond to the three segments, more clearly distinguished after processing the data, as shown later. However, a few samples from the centre parts of these larger zones display low values of the melting tracers, and some high values are seen near the discontinuities.

To reduce the impact of these extreme values and display larger-scale tendencies, we apply a seven-point running average to the data for each parameter, following the procedure already used by Langmuir *et al.*⁴ to characterize the along-ridge axis chemical variations of the oceanic lavas. After processing (Fig. 5), each parameter still displays variations, with the main peaks and dips located at positions close to those defined using the raw data, but with a smaller amplitude and a relatively longer wavelength because the number of chemical signals is reduced by 30% by the averaging method. This shows that the processed patterns represent the pre-smoothed data well, and that the data represent a geological reality.

To calculate the degree of conformity existing between the positions of peaks and valleys for the two melting parameters after processing, we defined a reference grid made by vertical lines that respectively locate the position of each peak (full line) and valleys displayed by the Cr/(Cr + Al) index. With this grid, we calculated that 95% of the peaks and valleys are common to the Cr/(Cr + Al) in spinel and the Yb/Er ratio in the whole rock, which indicates that these two partial melting tracers are perfectly correlated. The grid also allows us to calculate a correlation index of 61 to 70% between the processed and raw data that reinforces the idea that our modelling is correct. On the other hand, Fig. 5a and b also shows that the processed data allow us to define better the three large segments tentatively described by the raw data. The raw data indeed display large-scale gaussian-type distributions, with marked minima at points 40, 190, 330 and 400 km for both parameters, which are best-fitted using a third-degree polynomial equation. The data clearly varies with a large wavelength.

Implications for ridge segmentation

Our petrological data on the mantle residues show that, when considered as a whole, the Oman fossil oceanic mantle exhibits great chemical variability in terms of its partial melting history, because it underwent between 10 and 30% of melt extraction, as previously estimated for the southern massifs²⁴. But this large variability is well organized when considering the precise location of the samples along the inferred palaeo-ridge. To show this, we have examined the Cr/(Cr + Al) ratio in spinel and the Yb/Er ratio of the whole rock; this data is in agreement with our data on other partial melting parameters such as the Mg/(Mg + Fe) ratio in olivine, the Cr/(Cr + Al) in opx or the MgO and Al₂O₃ in the whole rock.

When plotted along-ridge, the partial melting tracers display two types of variations: one with a long wavelength and the other with a

shorter one. The long-wavelength variation defines large segments, 100 to 160 km long. Their extremities are marked by peridotites that bear the lowest values of the partial melting parameters, in contrast to what is observed in the central parts where the average values are the highest (Fig. 5c). Following ref. 4, the discontinuities should correspond to the zones of lowest magma production, which is in relatively good agreement with the Oman crust-thickness estimates¹³. They also roughly correspond to the zones of dyke penetrations (PZ on Fig. 5c) thought to reflect the ends of main magmatic cells¹⁰. We therefore conclude that the large segments defined in this study relate to the upwelling of large asthenospheric domes, as was already proposed for the origin of similar segments defined in the oceanic mantle on the basis of the lava compositions⁴. The shorter, 10–20-km-long segments, marked by pronounced variations of the chemical tracers, may relate the uplift of more local and superficial diapirs, similar to those described in the Oman ophiolite^{13,15}. They probably underline the feeding zones for the magmatic crust. According to the described variations, the mantle flow had a three-dimensional structure at relatively shallow levels along the Oman ridge, which indicates that the mantle had a relatively low viscosity, as expected for partially melted rocks⁴¹.

The lengths of the larger segments (100 to 160 km) mean that they cannot be considered as first-order segments as defined in ref. 2, but may correspond to second-order discontinuities such as OSC or NTD. Their spacing is within the range of the spacing calculated for half-spreading rates below 6 to 10 cm yr⁻¹ using recent numerical models that have a non-uniform mantle rheology including the effects of partial melting and melt⁴¹. On the other hand, the smallest, 10–20-km-long segments can be interpreted as third-order discontinuities such as small OSC or zero offset transform faults. At the ophiolite scale, the northern and southern extremities of our large segments are not chemically constrained, because it was impossible to obtain samples in these areas. They could, however, represent the boundaries of a larger-scale first-order segmentation if the extremities correspond to major transform faults⁴².

Our petrological data indicate that the peridotites locally suffered a very high degree of partial melting (it reaches 25–30% locally), as already proposed for some other Oman peridotites^{23,24}. This explains the local occurrence in the ophiolite of ultra-depleted magmas^{17,43–46}, which are generally thought to be formed in subduction zone contexts^{22,37}. But such magmas can also be related to the rise of hot mantle upwellings in the previously hydrated and depleted shallow mantle of the oceanic lithosphere¹⁷, which would be in better agreement with our oxygen fugacity estimates (which are those of an oceanic reduced mantle³⁶).

Although the segmentation features we have described in the Oman ophiolite cannot be used to discuss the ridge tectonic setting—large oceanic domain versus small back-arc basin—it does reinforce former interpretations indicating that this ophiolite formed at an oceanic ridge, probably intermediate to fast-spreading. □

Received 27 February; accepted 1 October 2004; doi:10.1038/nature03075.

1. Sempéré, J. C. & MacDonald, K. C. Marine tectonics: processes at mid-ocean ridges. *Rev. Geophys.* **25**, 1313–1347 (1987).
2. MacDonald, K. C., Fox, P. J. & Perram, L. J. A new view of the mid-ocean ridge from the behaviour of ridge-axis discontinuities. *Nature* **335**, 217–225 (1988).
3. Sinton, J. M., Smaglik, S. M., Mahoney, J. J. & MacDonald, K. C. Magmatic processes at superfast spreading mid-ocean ridges: glass compositional variations along the East Pacific Rise, 13°–23°S. *J. Geophys. Res.* **96**, 6133–6155 (1991).
4. Langmuir, C. H., Klein, E. M. & Planck, T. Petrological systematics of mid-ocean ridge basalts: constraints on melt generation beneath ocean ridges. *Geophys. Monogr. Ser.* **71**, 183–280 (1992).
5. Cannat, M. How thick is the magmatic crust at slow spreading ridges? *J. Geophys. Res.* **101**, 2847–2857 (1996).
6. Meibom, A. & Anderson, D. L. The statistical upper mantle assemblage. *Earth Planet. Sci. Lett.* **217**, 123–139 (2004).
7. Niu, Y., Bideau, D., Hekinian, R. & Batiza, R. Mantle compositional control on the extent of mantle melting, crust production, gravity anomaly, ridge morphology, and ridge segmentation: a case study at the Mid-Atlantic Ridge 33–35 degree N. *Earth Planet. Sci. Lett.* **186**, 383–399 (2001).
8. Hofmann, A. W. Mantle geochemistry: the message from oceanic volcanism. *Nature* **385**, 219–229 (1997).

9. Nicolas, A., Boudier, F., Ildefonse, B. & Ball, E. Accretion of Oman and United Arab Emirates Ophiolites—Discussion of a new structural map. *Mar. Geophys. Res.* **21**, 147–179 (2000).
10. MacLeod, C. J. & Rothery, D. A. Ridge axial segmentation in the Oman ophiolite: evidence from along-strike variations in the sheeted dyke complex. *Geol. Soc. Spec. Publ.* **60**, 39–63 (1992).
11. Reuber, I. Complexity of the crustal sequence in the northern Oman ophiolite (Fizh and southern Aswad blocks): the effect of early slicing? *Tectonophysics* **151**, 137–165 (1988).
12. Juteau, T., Beurier, M., Dahl, R. & Nehlig, P. Segmentation at a fossil spreading axis: the plutonic sequence of the Wadi Haymiliyah area (Haylajn Block, Sumail Nappe, Oman). *Tectonophysics* **151**, 167–197 (1988).
13. Nicolas, A. & Boudier, F. Large mantle upwellings and related variations in crustal thickness in the Oman ophiolite. *Geol. Soc. Am. Spec. Publ.* **349**, 67–74 (2000).
14. Nicolas, A., Ceuleneer, G., Boudier, F. & Misseri, M. Structural mapping in the Oman ophiolites: Mantle diapirism along an oceanic ridge. *Tectonophysics* **151**, 27–56 (1988).
15. Ceuleneer, G., Nicolas, A. & Boudier, F. Mantle flow patterns at an oceanic spreading centre: the Oman peridotites record. *Tectonophysics* **151**, 1–26 (1988).
16. Girardeau, J., Monnier, C., Launeau, P. & Quatrevaux, F. Kinematics of mantle flow beneath a fossil overlapping spreading center: The Wuqbah massif case, Oman ophiolite. *Geochem. Geophys. Geosyst.* **3**, 1–20 (2002).
17. Python, G. & Ceuleneer, G. Nature and distribution of dykes and related melt migration structures in the mantle section of the Oman ophiolite. *Geochem. Geophys. Geosyst.* **4** (7), 1–34, doi:10.1029/2002GC000354 (2003).
18. Takazawa, E., Okayasu, T. & Satoh, K. Geochemistry and origin of the basal lherzolites from the northern Oman ophiolite (northern Fizh block). *Geochem. Geophys. Geosyst.* **4**, 1–31 (2003).
19. Hellebrand, E., Snow, J. E. & Muhe, R. Mantle melting beneath Gakkel Ridge (Arctic Ocean): abyssal peridotite spinel compositions. *Chem. Geol.* **182**, 227–235 (2002).
20. Mercier, J. C. & Nicolas, A. Textures and fabrics of upper-mantle peridotites as illustrated by xenoliths from basalts. *J. Petrol.* **16**, 454–487 (1975).
21. Dick, H. J. B. & Bullen, T. Chromian spinel as a petrogenetic indicator in abyssal and alpine-type peridotites and spatially associated lavas. *Contrib. Mineral. Petrol.* **86**, 54–76 (1984).
22. Pearce, J. A., Barker, P. F., Edwards, S. J., Parkinson, I. J. & Leat, P. T. Geochemistry and tectonic significance of peridotites from the South Sandwich arc-basin system, South Atlantic. *Contrib. Mineral. Petrol.* **139**, 36–53 (2000).
23. Girardeau, J., Monnier, C., Le Mée, L. & Quatrevaux, F. The Wuqbah peridotite, central Oman ophiolite: petrological characteristics of the mantle in a fossil overlapping ridge setting. *Mar. Geophys. Res.* **23**, 43–56 (2002).
24. Godard, M., Jousellin, D. & Bodinier, J.-L. Relationships between geochemistry and structure beneath paleo-spreading centre: a study of the mantle section in the Oman ophiolite. *Earth Planet. Sci. Lett.* **180**, 133–148 (2000).
25. Niu, Y. & Hékinian, R. Basaltic liquids and harzburgitic residues in the Garrett Transform: a case study at fast-spreading ridges. *Earth Planet. Sci. Lett.* **146**, 243–258 (1997).
26. Batanova, V. G., Suhr, G. & Sobolev, A. V. Origin of geochemical heterogeneity in the mantle peridotites from the Bay of Islands ophiolite, Newfoundland, Canada: ion probe study of clinopyroxenes. *Geochim. Cosmochim. Acta* **62**, 853–866 (1998).
27. Hellebrand, E., Snow, J. E. & Dick, H. J. B. Coupled major and trace elements as indicators of the extent of melting in mid-ocean-ridge peridotites. *Nature* **410**, 677–681 (2001).
28. Bodinier, J.-L., Vasseur, G. & Vernières, J. Mechanisms of mantle metasomatism: geochemical evidence from the Lherz Orogenic peridotite. *J. Petrol.* **31**, 597–628 (1990).
29. Vernières, J., Godard, M. & Bodinier, J.-L. A plate model for the simulation of trace element fractionation during partial melting and magma transport in the earth's upper mantle. *J. Geophys. Res.* **102**, 24771–24784 (1997).
30. Gruau, G., Bernard-Griffiths, J. & Lecuyer, C. The origin of U-shaped rare earth patterns in ophiolite peridotites: assessing the role of secondary alteration and melt/rock reaction. *Geochim. Cosmochim. Acta* **62**, 3545–3560 (1998).
31. Godard, M., Bodinier, J.-L. & Vasseur, G. Effects of mineralogical reactions on trace element redistributions in mantle rocks during percolation processes: A chromatographic approach. *Earth Planet. Sci. Lett.* **133**, 449–461 (1995).
32. Kelemen, P. B., Johnson, K. T. M., Kinzler, R. J. & Irving, A. J. High-field-strength element depletions in arc basalts due to mantle: production of harzburgites by mantle–magma interaction. *Nature* **358**, 635–640 (1990).
33. Johnson, K. M., Dick, H. J. B. & Shimizu, N. Melting in the oceanic upper mantle: an ion microprobe study of diopsides in abyssal peridotites. *J. Geophys. Res.* **95**, 2661–2678 (1990).
34. Niu, Y. Mantle melting and melt extraction processes beneath ocean ridges: evidence from abyssal peridotites. *J. Petrol.* **38**, 1047–1074 (1997).
35. Ballhaus, C., Berry, R. F. & Green, D. H. Oxygen fugacity controls in the Earth's upper mantle. *Nature* **348**, 437–440 (1990).
36. Bryndzia, L. T. & Wood, B. J. Oxygen thermobarometry of abyssal spinel peridotites: the redox state and the C–O–H volatile composition of the earth's sub-oceanic mantle. *Am. J. Sci.* **290**, 1093–1116 (1990).
37. Parkinson, I. J. & Arculus, R. J. The redox state of subduction zones: insights from arc-peridotites. *Chem. Geol.* **160**, 409–423 (1999).
38. Klein, E. M. & Langmuir, C. H. Global correlation of mid-ocean ridge basalt chemistry with axial depth and crustal thickness. *J. Geophys. Res.* **92**, 8089–8115 (1987).
39. Bonatti, E. & Michael, P. J. Mantle peridotites from continental rifts to ocean basins to subduction zones. *Earth Planet. Sci. Lett.* **91**, 297–311 (1989).
40. Sack, R. O. & Ghiorso, M. S. Chromian spinel as a petrogenetic indicator: thermodynamics and petrological implications. *Am. Mineral.* **76**, 827–847 (1991).
41. Choblet, G. & Parmentier, E. M. Mantle upwelling and melting beneath slow spreading centers: effects of variable rheology and melt productivity. *Earth Planet. Sci. Lett.* **184**, 589–604 (2001).
42. Ceuleneer, G. *Structure of the Oman Ophiolite: Mantle Flow Pattern Below an Oceanic Spreading Center and Thrusting at an Oceanic Ridge*. 338, PhD thesis, Univ. Nantes (1986).
43. Amri, I., Benoit, M. & Ceuleneer, G. Tectonic setting for the genesis of oceanic plagiogranites: evidence from a paleo-spreading structure in the Oman ophiolite. *Earth Planet. Sci. Lett.* **139**, 177–194 (1996).
44. Alabaster, T., Pearce, J. A. & Nalpas, J. The volcanic stratigraphy and petrogenesis of the Oman ophiolite complex. *Contrib. Mineral. Petrol.* **81**, 168–183 (1982).
45. Ernewein, M., Plüm, C. & Whitechurch, H. The death of an accretion zone as evidenced by the magmatic history of the Sumail ophiolite (Oman). *Tectonophysics* **151**, 247–274 (1988).
46. Smewing, J. D. Mixing characteristics and compositional differences in mantle-derived melts beneath spreading axes: Evidence from cyclically layered rocks in the ophiolite of North Oman. *J. Geophys. Res.* **86**, 2645–2659 (1981).
47. Barnes, S. J. & Roeder, P. L. The range of spinel compositions in terrestrial mafic and ultramafic rocks. *J. Petrol.* **42**, 2279–2302 (2001).
48. Ishii, T., Robinson, P. T., Maekawa, H. & Fiske, R. Petrological studies of peridotites from diapiric serpentinite seamounts in the Izu-Ogasawara-Mariana forearc, Leg 125. *Proc. ODP Sci. Res.* **125**, 445–485 (1992).
49. Monnier, C., Girardeau, J., Maury, R. & Cotten, J. Back-arc basin origin for the East Sulawesi ophiolite (eastern Indonesia). *Geology* **23**, 851–854 (1995).
50. Sun, S. S. & McDonough, W. F. Chemical and isotopic systematics of oceanic basalts: implications for mantle composition and processes. *Geol. Soc. Lond. Spec. Publ.* **42**, 313–345 (1989).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank G. Ceuleneer for his help during this study. We are indebted to H. Al Azri and to J.-P. Breton for their support. We are grateful to M. Polvé who let us work in her geochemistry laboratory in Toulouse. We also thank B. Reynier, M. Valladon and J. Cotten for their help during ICP-MS and ICP-AES analyses, and M. Bohn who is in charge of the electron microprobe. We also thank E. Boeuf and H. Loyer for help with the samples and thin-section preparations. Financial support was provided by the Centre National de la Recherche Scientifique (INSU-SDU) through the IT programme and by the MEN (to L.L.M.).

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to L.L.M. (Laurent.Lemee@chimie.univ-nantes.fr).

Therapeutic silencing of an endogenous gene by systemic administration of modified siRNAs

Jürgen Soutschek¹, Akin Akinc², Birgit Bramlage¹, Klaus Charisse², Rainer Constien¹, Mary Donoghue², Sayda Elbashir², Anke Geick¹, Philipp Hadwiger¹, Jens Harborth², Matthias John¹, Venkatasamy Kesavan², Gary Lavine², Rajendra K. Pandey², Timothy Racie², Kallanthottathil G. Rajeev², Ingo Röhl¹, Ivanka Toudjarska², Gang Wang², Silvio Wuschko¹, David Bumcrot², Victor Kotliansky², Stefan Limmer¹, Muthiah Manoharan² & Hans-Peter Vornlocher¹

¹Alnylam Europe AG, Fritz-Hornschuch-Str. 9, 95326 Kulmbach, Germany

²Alnylam Pharmaceuticals Inc., 300 3rd Street, Cambridge, Massachusetts 02142, USA

RNA interference (RNAi) holds considerable promise as a therapeutic approach to silence disease-causing genes, particularly those that encode so-called 'non-druggable' targets that are not amenable to conventional therapeutics such as small molecules, proteins, or monoclonal antibodies. The main obstacle to achieving *in vivo* gene silencing by RNAi technologies is delivery. Here we show that chemically modified short interfering RNAs (siRNAs) can silence an endogenous gene encoding apolipoprotein B (apoB) after intravenous injection in mice. Administration of chemically modified siRNAs resulted in silencing of the apoB messenger RNA in liver and jejunum, decreased plasma levels of apoB protein, and reduced total cholesterol. We also show that these siRNAs can silence human apoB in a transgenic mouse model. In our *in vivo* study, the mechanism of action for the siRNAs was proven to occur through RNAi-mediated mRNA degradation, and we determined that cleavage of the apoB mRNA occurred specifically at the predicted site. These findings demonstrate the therapeutic potential of siRNAs for the treatment of disease.

RNAi has been applied widely as a target validation tool in post-genomic research, and it represents a potential strategy for *in vivo* target validation and therapeutic product development¹. *In vivo* gene silencing with RNAi has been reported using both viral vector delivery² and high-pressure, high-volume intravenous (i.v.) injection of synthetic siRNAs³, but these approaches have limited if any clinical use. *In vivo* gene silencing has also been reported after local, direct administration (intravitreal, intranasal and intrathecal) of siRNAs to sequestered anatomical sites in models of choroidal neovascularization⁴, lung ischaemia-reperfusion injury⁵ and neuropathic pain⁶, respectively. These reported approaches demonstrate the potential for delivery to organs such as the eye, lungs and central nervous system. However, there are no published reports of systemic activity for siRNAs towards endogenous targets after conventional and clinically acceptable routes of administration. A critical requirement for achieving systemic RNAi *in vivo* is the introduction of 'drug-like' properties, such as stability, cellular delivery and tissue bioavailability, into synthetic siRNAs.

Conferring drug-like properties on siRNAs

In exploring the potential of synthetic siRNAs to silence endogenous target genes, we found that chemically stabilized and cholesterol-conjugated siRNAs⁷ have markedly improved pharmacological properties *in vitro* and *in vivo*. Chemically stabilized siRNAs with partial phosphorothioate backbone and 2'-O-methyl sugar modifications on the sense and antisense strands showed significantly enhanced resistance towards degradation by exo- and endonucleases in serum and in tissue homogenates. The conjugation of cholesterol to the 3' end of the sense strand of a siRNA molecule by means of a pyrrolidine linker (thereby generating chol-siRNA) did not result in a significant loss of gene-silencing activity in cell culture. Furthermore, unlike unconjugated siRNAs, a chol-siRNA directed to luciferase (chol-luc-siRNA) showed reduction in luciferase activity in HeLa cells transiently expressing luciferase, with a half-maximal inhibitory concentration (IC₅₀) of about 200 nM in

the absence of transfection reagents or electroporation.

Binding of chol-siRNAs to human serum albumin (HSA) was determined by surface plasmon resonance measurement (data not shown). Unconjugated siRNAs demonstrated no measurable binding to HSA, whereas chol-siRNAs bound to HSA with an estimated dissociation constant (K_d) of 1 μ M. Presumably because of enhanced binding to serum proteins, chol-siRNAs administered to rats by i.v. injection showed improved *in vivo* pharmacokinetic properties as compared to unconjugated siRNAs. After i.v. injection in rats at 50 mg kg⁻¹, radioactively labelled chol-siRNAs had an elimination half life (two compartments), $t_{1/2}$ of 95 min and a corresponding plasma clearance (C_L) of 0.5 ml min⁻¹, whereas unconjugated siRNAs had a $t_{1/2}$ of 6 min and C_L of 17.6 ml min⁻¹. As measured by an RNase protection assay (RPA), chol-siRNAs showed broad tissue biodistribution 24 h after injection in mice. Although no detectable amounts of unconjugated siRNAs were observed in tissue samples, significant levels of chol-siRNAs were detected in liver, heart, kidney, adipose, and lung tissue samples. Together, these studies demonstrate that cholesterol conjugation significantly improves *in vivo* pharmacological properties of siRNAs.

Selection of apoB as an endogenous gene target

Apolipoprotein B is the essential protein for formation of low-density lipoproteins (LDL) in metabolism of dietary and endogenous cholesterol, and is the ligand for the LDL receptor⁸. Mouse apoB is a large protein of 4,515 amino acids and is expressed predominantly in liver and jejunum. apoB mRNA is subject to post-transcriptional editing, and the unedited and edited transcripts encode the full-length protein apoB-100, and a carboxy-terminal truncated isoform, apoB-48, respectively. In mice, editing of apoB mRNA occurs in both the liver and jejunum: apoB-48 is the predominant protein form in the jejunum and both apoB-48 and apoB-100 are expressed in the liver. Heterozygous knockout mice for apoB show a 20% decrease in cholesterol levels and are resistant to

diet-induced hypercholesterolaemia⁹.

Serum levels of apoB, LDL and cholesterol correlate significantly with increased risk of coronary artery disease (CAD). A diminished number of functional LDL receptors on the cell surface, disrupting receptor-mediated removal of apoB-containing LDL from circulation, has been identified as the basis for familial hypercholesterolaemia (FH)¹⁰. Patients with homozygous and heterozygous FH have accelerated CAD leading to premature atherosclerosis and cardiac mortality. Conversely, patients with hypobetalipoproteinaemia have reduced levels of LDL and cholesterol and are at reduced risk for CAD¹¹. Accordingly, lowering of serum cholesterol and LDL levels is a predominant clinical strategy for management of CAD and is achieved by modification of dietary sources of cholesterol and/or inhibition of endogenous cholesterol synthesis with pharmacological therapies. Notwithstanding significant improvements in the management of CAD with these approaches, millions of patients remain at significant risk for CAD and its clinical sequelae—acute coronary syndromes such as myocardial infarction and cardiac mortality—due to advanced atherosclerosis from intractably high levels of cholesterol and LDL. Clearly, new therapeutic strategies are needed. Accordingly, apoB, a protein not amenable to inhibition by conventional small-molecule- or protein-based therapeutics, was selected as a potential clinical target for development of siRNA therapeutics.

Using conventional bioinformatics, 84 siRNAs specific for both human and mouse apoB mRNA were designed and synthesized (data not shown). These apoB-siRNAs were screened for their ability to reduce apoB mRNA and protein levels, as measured by polymerase chain reaction with reverse transcription (RT-PCR) and enzyme-linked immunosorbent assay (ELISA), respectively, in HepG2 liver cells after transfection at a concentration of 100 nM. Five apoB-siRNAs were identified that reduced both mRNA and protein levels by >70%. Because exonucleolytic degradation is the predominant mechanism for siRNA degradation in serum, two selected apoB-siRNAs (apoB-1-siRNA and apoB-2-siRNA) and one four-nucleotide mismatch control for apoB-1-siRNA (mismatch-siRNA) were stabilized at the 3' end of the sense and antisense strands by phosphorothioate backbone modifications and additional incorporation of two 2'-O-methyl nucleotides at the 3' end of the antisense strand. Chol-siRNAs were synthesized by linkage of cholesterol to the 3' end of the sense strand via a pyrrolidine linker. Chol-apoB-1-siRNA was significantly more stable than unconjugated apoB-1-siRNA in human serum: gel electrophoresis showed >50% intact chol-apoB-1-siRNA after a 1 h incubation at 37 °C compared with <5% intact unconjugated apoB-1-siRNA. Similar data were obtained for chol-apoB-2-siRNA, although this siRNA was less stable than chol-apoB-1-siRNA. Dose

response curves for the activity of conjugated and unconjugated apoB-specific and control siRNAs were measured in HepG2 cells using transfection. Two conjugated control siRNAs (chol-luc-siRNA and chol-mismatch-siRNA) showed no significant inhibition of apoB protein expression at concentrations as high as 30 nM. In contrast, three specific siRNAs (unconjugated apoB-1-siRNA, chol-apoB-1-siRNA and chol-apoB-2-siRNA) showed dose-dependent silencing of apoB protein expression based on apoB ELISA measurements—IC₅₀ values of 0.5 nM, 5 nM and 8 nM were calculated, respectively.

In vivo studies with modified siRNAs

To demonstrate the ability of chol-apoB-siRNAs to silence apoB expression *in vivo*, experiments were first performed in C57BL/6

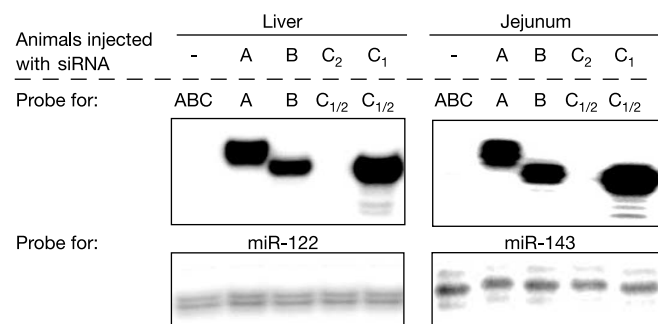


Figure 1 Biodistribution of siRNAs in liver and jejunum. An RPA was used to detect siRNAs in pooled liver and jejunum tissue lysates from animals injected with saline (–), chol-luc-siRNA (A), chol-mismatch-siRNA (B), unconjugated apoB-1-siRNA (C₂) or chol-apoB-1-siRNA (C₁). Detection by RPA of endogenous miRNAs in liver (miR-122) and jejunum (miR-143) served as an internal loading control.

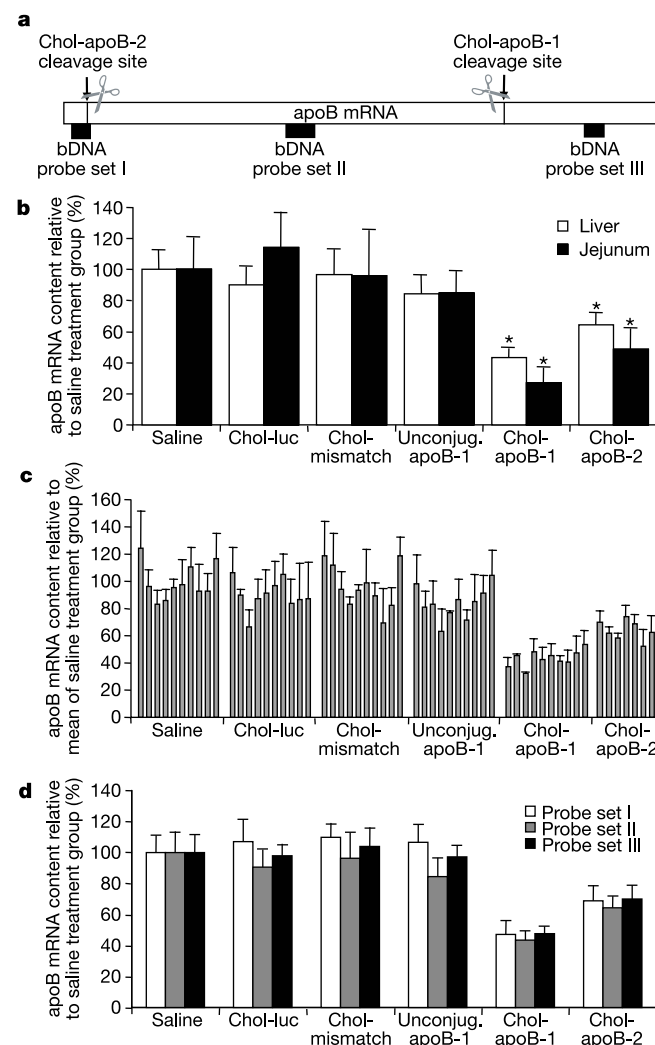


Figure 2 *In vivo* silencing of murine apoB mRNA by siRNAs in wild-type mice. Treatment groups comprised saline control ($n = 10$), chol-luc-siRNA control ($n = 10$), chol-mismatch-siRNA control ($n = 10$), unconjugated apoB-1-siRNA ($n = 10$), chol-apoB-1-siRNA ($n = 10$) and chol-apoB-2-siRNA ($n = 7$). bDNA measurements were performed with probe set II. Error bars represent the standard deviation (s.d.) of the mean. Statistical analysis was by analysis of variance (ANOVA) with Bonferroni post-hoc *t*-test, one-tailed. Asterisk, $P < 0.0001$ compared with saline control animals. **a**, Schematic representation of the apoB mRNA illustrating the binding regions of three bDNA probe sets in relation to the two siRNA cleavage sites. **b**, Effects of siRNA administration on mean apoB mRNA levels. **c**, apoB mRNA levels from individual mice treated with saline or siRNAs. Data are mean values from three liver samples from each individual animal. **d**, Effects of siRNA administration on the reduction of apoB mRNA measured by bDNA assays using three different probe sets.

mice fed a normal chow diet. siRNAs were administered by tail-vein injection with normal volume (0.2 ml) and normal pressure. Biodistribution of siRNAs was assessed by RPA of siRNAs in tissue samples from liver and jejunum obtained 24 h after the last injection. Significant levels of chol-luc-siRNA, chol-apoB-1-siRNA and chol-mismatch-siRNA were detected in liver and jejunum ($100\text{--}200\text{ ng g}^{-1}$ tissue for chol-apoB-1-siRNA), whereas levels of unconjugated apoB-1-siRNA were below our detection limit (Fig. 1). Levels of chol-apoB-2-siRNA were also detected but at levels approximately 10% of those observed for other chol-siRNAs.

The primary measure of RNAi-mediated effects is the reduction (that is, silencing) of the target mRNA. To measure silencing of apoB mRNA, we used a branched-DNA (bDNA) detection method and bDNA probes (Fig. 2a) to quantify apoB mRNA levels in liver and jejunum, two organs where apoB is known to be expressed. As shown in Fig. 2b, mice treated with chol-apoB-1-siRNA and chol-apoB-2-siRNA showed statistically significant reductions (mean $\pm$ s.d.; $57 \pm 6\%$ and $36 \pm 8\%$, respectively) in apoB mRNA levels in liver samples as compared with saline control ($P < 0.0001$). In jejunum tissue samples, mice injected with chol-apoB-1-siRNA and chol-apoB-2-siRNA showed an even more substantial reduction in apoB mRNA levels of $73 \pm 10\%$ and $51 \pm 13\%$, respectively, as compared with saline control ($P < 0.0001$). Individual animal results for apoB mRNA reduction in the liver are shown in Fig. 2c and demonstrate the consistent and robust effect observed for specific chol-siRNAs as compared with other treatment groups. Similar results were observed for apoB mRNA reduction in the jejunum from individual animals (data not shown). Owing to the extended length of the apoB mRNA, two additional probes at the distal ends of the apoB open reading frame (ORF) were designed. As measured with the three divergent probe sets, identical levels of apoB mRNA reduction were detected for animals treated with chol-apoB-1-siRNA and chol-apoB-2-siRNA (Fig. 2d). These data suggest a uniform and rapid degradation of apoB mRNA after treatment with chol-apoB-siRNAs, and argue against the potential existence of truncated amino-terminal apoB protein fragments translated from incompletely degraded siRNA-cleavage products, as has been reported for ribozyme-mediated cleavage of apoB mRNA¹².

Silencing of the apoB mRNA would be expected to result in a corresponding reduction in apoB protein levels. An ELISA-based method specific for detection of apoB-100 protein was used to measure the effects of chol-apoB-siRNA treatment on plasma levels of apoB protein. In addition to the effects on apoB mRNA levels, treatment with chol-apoB-1-siRNA and chol-apoB-2-siRNA reduced plasma levels of apoB-100 protein 24 h after siRNA treatment by $68 \pm 14\%$ and $31 \pm 18\%$, respectively, compared with

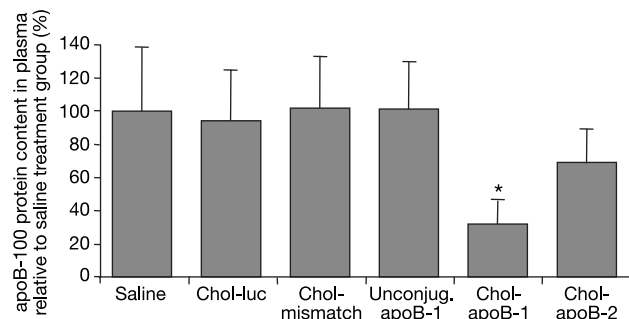


Figure 3 Effects of siRNA administration on apoB-100 protein levels. Average plasma levels of apoB-100 protein for the different treatment groups as measured by ELISA. Error bars represent the s.d. of the mean. Statistical analysis was by ANOVA with Bonferroni post-hoc *t*-test, one-tailed. Asterisk, $P < 0.0001$ compared with saline control animals.

levels in saline-treated control animals (Fig. 3). These results achieved statistical significance ($P < 0.0001$) for the group treated with the more potent and stable chol-apoB-1-siRNA. As the LF3 antibody used in this study recognizes only apoB-100, and not apoB-48, the observed apoB-100 reduction may underestimate the full effect of chol-apoB-1-siRNA at the protein level.

To confirm the physiological relevance of apoB mRNA silencing on lipoprotein metabolism, we characterized the effect of siRNA treatment and the resulting reduction of apoB protein levels on lipoprotein profiles and cholesterol levels. Using an NMR-based method, complete lipoprotein profiles were generated and concentrations of chylomicrons, very-low-density lipoprotein (VLDL), LDL and high-density lipoprotein (HDL) particles were calculated (Fig. 4a). As expected, HDL represented the predominant lipoprotein fraction in mouse plasma. Similar to results observed in heterozygous knockout mice for apoB⁹, treatment with chol-apoB-1-siRNA resulted in a 25% reduction in HDL particle

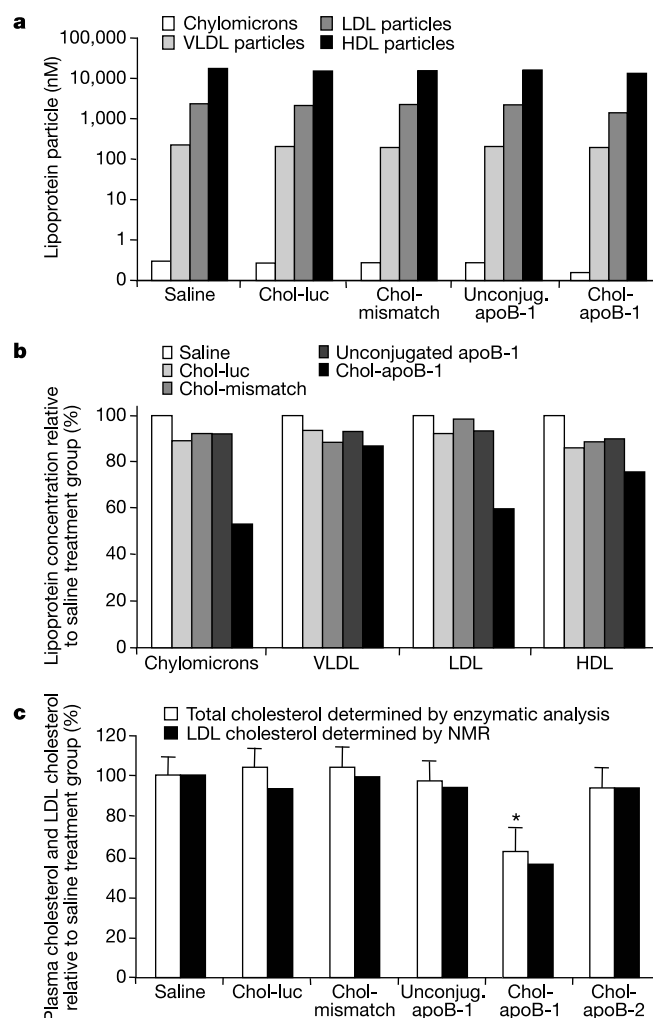


Figure 4 Therapeutic reduction of lipoprotein and cholesterol levels after siRNA treatment. **a**, Lipoprotein profile of pooled plasma samples from treatment groups determined by NMR analysis. **b**, Relative reduction of lipoprotein classes for the siRNA treatment groups normalized against the average levels of saline control group. **c**, Effects of siRNA administration on plasma cholesterol and LDL cholesterol. Plasma cholesterol was determined by enzymatic assay and LDL cholesterol calculated from NMR measurements. Error bars represent the s.d. of the mean. Statistical analysis was by ANOVA with Bonferroni post-hoc *t*-test, one-tailed. Asterisk, $P < 0.0001$ compared with saline and chol-mismatch-siRNA control animals. NMR data are based on single measurements of pooled plasma from treatment groups.

concentration (Fig. 4b). Furthermore, treatment of mice with chol-apoB-1 siRNA resulted in an almost 50% reduction of chylomicron levels and an approximately 40% reduction in LDL levels, whereas VLDL levels were not altered. Treatment with either of the control siRNAs did not change the lipoprotein profile significantly. In addition to reductions in lipoprotein concentrations, *in vivo* silencing of apoB by chol-apoB-1-siRNA led to a significant reduction ($37 \pm 11\%$; $P < 0.0001$) of total plasma cholesterol as compared with saline control animals (Fig. 4c). Treatment with the less potent chol-apoB-2-siRNA failed to show significant reductions in cholesterol, consistent with the reduced activity of this chol-siRNA on apoB mRNA and protein levels. Treatment with chol-apoB-1-siRNA also resulted in a 44% decrease in LDL-associated cholesterol, consistent with the effects observed on apoB protein levels. In aggregate, the effects on cholesterol reduction and lipoprotein profiles would be considered highly clinically significant in patients with hypercholesterolaemia, and actually exceed the level of cholesterol reduction observed in heterozygous apoB knockout mice⁹.

To extend our findings of *in vivo* silencing by chol-apoB-siRNAs in normal mice, we performed an additional study in a human apoB transgenic mouse model¹³. These mice express human apoB-100 in liver and have elevated levels of apoB as compared with normal mice; when fed a high-fat diet, these mice develop severe atherosclerosis¹⁴. In our experiments, we administered saline, chol-mismatch-siRNA and chol-apoB-1-siRNA to apoB transgenic mice fed a normal chow diet. As shown in Fig. 5, chol-apoB-1-siRNA brought about a significant reduction of endogenous murine apoB expressed in both liver and jejunum tissue samples ($P < 0.0001$, relative to saline and chol-mismatch-siRNA treatment). Relative to the saline control, levels of murine apoB mRNA were reduced by $57 \pm 10\%$ in liver and $42 \pm 12\%$ in jejunum. In addition, chol-apoB-1-siRNA, which was selected in part owing to its sequence identity to both human and mouse apoB, showed significant silencing of the human transgene expressed in the liver, where human apoB mRNA was silenced by $60 \pm 10\%$ ($P < 0.0001$). In contrast to these effects, chol-mismatch-siRNA showed no effect on mouse or human apoB mRNA levels. These results confirm the effect of specific chol-siRNAs on apoB silencing in a different mouse model. Moreover, this specific chol-siRNA was shown to silence a transgenic human mRNA *in vivo*.

An important consideration for siRNA-mediated inhibition of gene expression is whether the observed effects are specific and not due to nonspecific "off target" effects¹⁵ and potential interferon responses¹⁶, which have been reported with siRNAs *in vitro* and other oligonucleotide-based approaches *in vivo*. In our experiments, the effects of apoB-specific, cholesterol-conjugated siRNAs were seen with two divergent siRNAs targeting separate sequence regions of the apoB mRNA. Furthermore, the *in vivo* silencing of

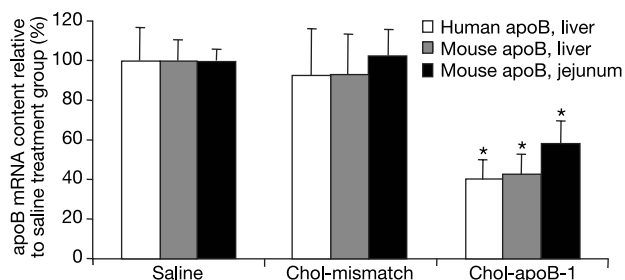


Figure 5 *In vivo* silencing of murine and human apoB mRNA in mice transgenic for human apoB. Reduction of human and mouse apoB mRNA levels in mice transgenic for human apoB that received saline ($n = 8$), chol-mismatch-siRNA ($n = 8$) and chol-apoB-1-siRNA ($n = 8$). Statistical analysis was by ANOVA with Bonferroni post-hoc *t*-test, one-tailed. Asterisk, $P < 0.0001$ compared with saline and chol-mismatch-siRNA control animals. Error bars illustrate s.d. of the mean.

apoB by these siRNAs was specific as neither an irrelevant siRNA (chol-luc-siRNA) nor a mismatch control siRNA (chol-mismatch-siRNA)—although present at comparable concentrations in liver and jejunum—mediated a significant reduction in apoB mRNA, plasma apoB protein levels, or total cholesterol. Finally, the silencing of apoB mRNA by chol-apoB-siRNAs in liver as measured by bDNA assay and normalization to GAPDH mRNA was also demonstrated with normalization to three other liver mRNAs, including factor VII, glucose-6-phosphatase and VEGF (Supplementary Fig. 1).

Determination of *in vivo* mechanism of action

To prove that the *in vivo* activity was due to siRNA-directed cleavage, we characterized specific mRNA cleavage products using a modified 5'-RACE (rapid amplification of cDNA ends) technique previously used to demonstrate microRNA (miRNA)-directed mRNA cleavage in plants¹⁷ and mouse embryos¹⁸. As it relates to the specific cleavage of apoB mRNA by apoB-1-siRNAs, total RNA from mice in the different treatment groups was isolated, and then PCR was used to reveal fragments of the predicted length in animals receiving chol-apoB-1-siRNA treatment (Fig. 6a). Identity of the PCR products was confirmed by direct sequencing of the excised bands, which demonstrated that cleavage occurred at the predicted position for the siRNA duplex. Indeed, sequencing revealed cleavage after position 10,061 of the apoB ORF, exactly ten nucleotides downstream of the 5' end of the siRNA antisense strand. Specific cleavage fragments were detected in both liver and jejunum of animals receiving chol-apoB-1-siRNA treatment (Fig. 6b). No fragments were detected in tissues of animals receiving control siRNAs (chol-luc-siRNA or chol-mismatch-siRNA) or saline. As expected, in this 5'-RACE experiment of apoB mRNA cleavage mediated by chol-apoB-1-siRNA, no fragments were detected in tissues from animals receiving the alternative apoB-specific siRNA (chol-apoB-2-siRNA). Notably, a low level of specific cleavage product was detected in the jejunum of animals receiving the unconjugated apoB-1-siRNA despite no evidence for significant knockdown of total apoB mRNA levels by this siRNA. This indicates

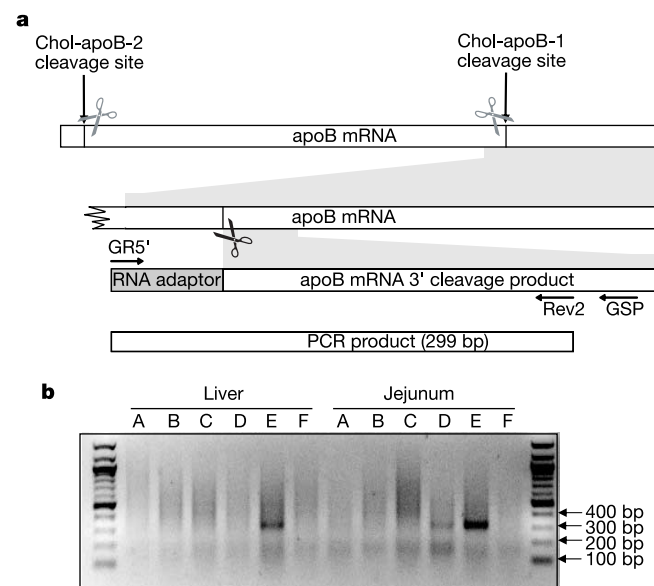


Figure 6 siRNA-mediated cleavage of apoB mRNA *in vivo*. **a**, Schematic representation of the apoB mRNA illustrating siRNA cleavage sites and RACE strategy to detect cleavage product. Cleaved mRNA ligated to an RNA adaptor was reverse transcribed using primer GSP. **b**, Agarose gel of 5'-RACE-PCR amplification, using the primer pair GR5' and Rev2, showing specific cleavage products in liver and jejunum. Treatment groups are: A, saline; B, chol-luc-siRNA; C, chol-mismatch-siRNA; D, apoB-1-siRNA; E, chol-apoB-1-siRNA; F, chol-apoB-2-siRNA.

that some unconjugated apoB-1-siRNA is able to enter epithelial cells of the jejunum after systemic administration despite lacking cholesterol conjugation. Together, these data demonstrate that inhibition of apoB was achieved by an RNAi mechanism of action. To our knowledge, this is the first demonstration of silencing of an endogenous gene in mammals by a mechanism of RNAi-mediated degradation of the target mRNA.

Discussion

Our findings demonstrate that RNAi can be used to silence endogenous genes involved in the cause or pathway of human disease with a clinically acceptable formulation and route of administration by means of systemic delivery. In our study, we have shown that the mechanism of action for chemically modified siRNAs was by RNAi-mediated degradation of the target mRNA. Chol-apoB-siRNAs, but not unconjugated apoB-siRNAs, showed biological activity, demonstrating an important role for cholesterol conjugation of siRNAs to achieve systemic *in vivo* activity, and suggesting the opportunity to further optimize systemic activity through chemical conjugation strategies. Indeed, further optimization is warranted to achieve improved *in vivo* potency for chol-siRNAs at doses and dose regimens that are clinically acceptable. Nevertheless, these findings hold promise for the development of a new class of therapeutics that harnesses the RNAi mechanism. Of particular interest is the use of RNAi therapeutics to silence genes (such as the apoB gene) or mutated or variant alleles whose proteins are refractory to the discovery of traditional small molecules or biotherapeutic drugs. □

Methods

Synthesis of siRNAs

The siRNAs used in this study consisted of a 21-nucleotide sense strand and a 23-nucleotide antisense strand resulting in a two-nucleotide overhang at the 3' end of the antisense strand. apoB-1-siRNA (ORF position 10049–10071): sense 5'-GUCACACAC UGAUACCAA*U-3', antisense 5'-AUUGGUAUUCAGUGAUGAC*a*C-3'; chol-apoB-1-siRNA: sense 5'-GUCACACACUAAUACCAA*chol-3', antisense 5'-AUUGGUAUUCAGUGAUGAC*a*C-3'; chol-mismatch-siRNA: sense 5'-GUGAUCAGACUAAUACCAA*chol-3', antisense 5'-AUUCGUAUUGAGUCUGA UCAc*a*C-3'; chol-apoB-2-siRNA (ORF position 327–349): sense 5'-AGGUGUAUGGC UUCAACCCUG*chol-3', antisense 5'-CAGGGUAGAAGCAUACACCu*c*U-3'; chol-luc-siRNA: sense 5'-GAACUGUGUGAGAGGUCCU*chol-3', antisense 5'-AGGAC CUCUCACACAGUUC*g*C-3'. The lower-case letters represent 2'-O-methyl-modified nucleotides; asterisks represent phosphorothioate linkages.

RNA oligonucleotides were synthesized using commercially available 5'-O-(4,4'-dimethoxytrityl)-3'-O-(2-cyanoethyl-N,N-diisopropyl) phosphoramidite monomers of uridine (U), 4-N-benzoylcytidine (C^{Bz}), 6-N-benzoyladenine (A^{Bz}) and 2-N-isobutrylguanosine (G^{ibu}) with 2'-O-t-butylidimethylsilyl protected phosphoramidites and the corresponding 2'-O-methyl phosphoramidites according to standard solid phase oligonucleotide synthesis protocols¹⁹. After cleavage and de-protection, RNA oligonucleotides were purified by anion-exchange high-performance liquid chromatography and characterized by ES mass spectrometry and capillary gel electrophoresis. RNA with phosphorothioate backbone at a given position was achieved by oxidation of phosphite with Beaucage reagent²⁰ during oligonucleotide synthesis. Chol-siRNAs were synthesized using the same protocols as above except that the RNA synthesis started from a controlled-pore glass solid support carrying a cholesterol-aminocaproic acid-pyrrolidine linker (V.Ke., K.G.R. and M.M., unpublished data). For this support, the first nucleotide linkage was achieved using a phosphorothioate linkage to provide additional 3'-exonuclease stability. To generate siRNAs from RNA single strands, equimolar amounts of complementary sense and antisense strands were mixed and annealed, and siRNAs were further characterized by native gel electrophoresis.

In vitro activity and stability assays

To determine *in vitro* activity of siRNAs, HepG2 cells were transfected with siRNAs using oligofectamine (Invitrogen) and siRNA concentrations ranging from 0.1, 0.3, 1, 3, 10 to 30 nM. apoB protein concentration was determined from cell culture supernatant by a sandwich ELISA capturing apoB with a polyclonal goat anti-human apoB antibody (Chemicon International). apoB detection was performed with a horseradish peroxidase-conjugated goat anti-human apoB-100 polyclonal antibody (Academy Bio-Medical Company). The remaining apoB protein content was calculated as the ratio of apoB protein concentration in the supernatant of cells treated with the apoB-specific siRNA duplex to the apoB concentration in the supernatant of cells treated with an unspecific control siRNA duplex. Mouse serum (Sigma-Aldrich Chemie GmbH) was used for stability assays. Double-stranded RNAs (5 µM) were incubated in 95% serum, and the mixture was incubated at 37 °C for various lengths of time (for example, 0, 15 or 30 min, or 1, 2, 4, 8, 16 or 24 h). siRNAs were isolated by hot phenol extraction in the presence of

sodium dodecyl sulphate followed by ethanol precipitation. Re-suspended RNA samples were run on a denaturing 14% polyacrylamide gel containing 20% formamide for 2 h at 45 mA. RNA bands were visualized by staining with the 'Stains-All' reagent (Sigma-Aldrich Chemie GmbH) according to the manufacturer's instructions.

In vivo silencing experiments

C57BL/6 mice received, on three consecutive days, tail vein injections of saline or different siRNAs. All siRNAs were administered at doses of 50 mg kg⁻¹ in approximately 0.2 ml per injection. Measurements of apoB mRNA, apoB protein levels, lipoprotein concentrations and plasma cholesterol content were performed 24 h after the last i.v. injection. Experiments were carried out in a blinded fashion. The same experimental design was used for experiments with the human apoB transgenic mice (1004-T hemizygotes, Taconic).

In vivo bioanalytical methods

An RPA, using radiolabelled probes complementary to the antisense strands, was used to detect siRNAs in pooled liver and jejunum tissue lysates from animals treated with saline or siRNAs. RPA for endogenous miRNAs was used as a loading control for jejunum (miR-143, sequence 5'-UGAGAUGAAGCACUGUAGCUCA-3') and liver (miR-122, 5'-UGGAGUGUGACAAUGGUGUUUG-3').

The QuantiGene assay (Genospectra) was used to quantify the reduction of mouse apoB mRNA in liver and jejunum tissue after siRNA treatment. Small uniform tissue samples where collected 24 h after the last injection. Lysates from three tissue samples per animal were directly used for apoB and GAPDH mRNA quantification, and the ratio of apoB and GAPDH mRNA was calculated and expressed as a group average relative to the saline control group. Specific probes for detection of apoB mRNA levels were designed to the following regions of the apoB mRNA ORF: probe set I 83–385; probe set II 5,045–5,673; probe set III 12,004–12,411. Furthermore, apoB mRNA reduction in liver was quantified from purified (RNeasy mRNA isolation kit, Qiagen), pooled mRNA for each treatment group. As well as GAPDH, factor VII, glucose-6-phosphatase and VEGF mRNAs were also used for normalization.

ELISA was used to quantify the reduction of apoB-100 protein levels in mouse plasma after siRNA treatment. apoB-100 from plasma samples of individual animals was detected using the primary antibody LF3 against mouse apoB-100 (gift of S. Young; see ref. ²¹). Levels were normalized to plasma volume and expressed as group averages relative to the saline control group.

Total cholesterol levels in the plasma were measured using the Cholesterol detection kit (Diasys). For NMR determination of the plasma lipoprotein profile a Bruker DRX 600 with cryoprobe head was used (LipoFIT Analytic GmbH). Single measurements of 500 µl mouse plasma (pooled from ten animals per treatment group) were performed. The lipoprotein subclass distribution was calculated from the NMR data by using computer algorithms that are based on human blood standards²². The particle number for lipoprotein classes was calculated based on the correlation of known particle size and composition with the experimentally determined NMR signal intensity. On the basis of this correlation, the cholesterol content in the LDL fraction was computed. The cholesterol values calculated from NMR data were confirmed by the presence of comparable levels of total cholesterol in plasma and HDL-cholesterol as determined by enzymatic assays.

5'-RACE analysis

Total RNA (5 µg) from pooled liver and jejunum samples from animals treated with different siRNAs was ligated to a GeneRacer adaptor (Invitrogen) without prior treatment. Ligated RNA was reverse transcribed using a gene-specific primer (GSP: 5'-CTCCTG TTGCAGTAGAGTGACGCT-3'). To detect cleavage products, PCR was performed using primers complementary to the RNA adaptor (GR5': 5'-CTCTAGAGCGACTGGAGCAGC AGGACACTA-3') and apoB mRNA (Rev2: 5'-ACGCGTCGACGTGGGAGCATGGAGT TGGCAGTTGTTTC-3'). Amplification products were resolved by agarose gel electrophoresis and visualized by ethidium bromide staining. The identity of specific PCR products was confirmed by sequencing of the excised bands.

Received 2 September; accepted 20 October 2004; doi:10.1038/nature03121.

- Novina, C. D. & Sharp, P. A. The RNAi revolution. *Nature* **430**, 161–164 (2004).
- Scherr, M., Battmer, K., Dallmann, I., Ganser, A. & Eder, M. Inhibition of GM-CSF receptor function by stable RNA interference in a NOD/SCID mouse hematopoietic stem cell transplantation model. *Oligonucleotides* **13**, 353–363 (2003).
- Song, E. *et al.* RNA interference targeting Fas protects mice from fulminant hepatitis. *Nature Med.* **9**, 347–351 (2003).
- Reich, S. J. *et al.* Small interfering RNA (siRNA) targeting VEGF effectively inhibits ocular neovascularization in a mouse model. *Mol. Vis.* **9**, 210–216 (2003).
- Zhang, X. *et al.* Small interfering RNA targeting heme oxygenase-1 enhances ischemia-reperfusion-induced lung apoptosis. *J. Biol. Chem.* **279**, 10677–10684 (2004).
- Dorn, G. *et al.* siRNA relieves chronic neuropathic pain. *Nucleic Acids Res.* **32**, e49 (2004).
- Lorenz, C., Hadwiger, P., John, M., Vornlocher, H.-P. & Unverzagt, C. Steroid and lipid conjugates of siRNAs to enhance cellular uptake and gene silencing in liver cells. *Bioorg. Med. Chem. Lett.* **14**, 4975–4977 (2004).
- Burnett, J. R. & Barrett, P. H. Apolipoprotein B metabolism: tracer kinetics, models, and metabolic studies. *Crit. Rev. Clin. Lab. Sci.* **39**, 89–137 (2002).
- Farese, R. V. Jr, Ruland, S. L., Flynn, L. M., Stokowski, R. P. & Young, S. G. Knockout of the mouse apolipoprotein B gene results in embryonic lethality in homozygotes and protection against diet-induced hypercholesterolemia in heterozygotes. *Proc. Natl Acad. Sci. USA* **92**, 1774–1778 (1995).
- Brown, M. S. & Goldstein, J. L. A receptor-mediated pathway for cholesterol homeostasis. *Science* **232**, 34–47 (1986).
- Glueck, C. J. *et al.* Prospective 10-year evaluation of hypobetalipoproteinemia in a cohort of 772 firefighters and cross-sectional evaluation of hypocholesterolemia in 1,479 men in the National Health and Nutrition Examination Survey I. *Metabolism* **46**, 625–633 (1997).

12. Enjoji, M., Wang, F., Nakamuta, M., Chan, L. & Teng, B. B. Hammerhead ribozyme as a therapeutic agent for hyperlipidemia: production of truncated apolipoprotein B and hypolipidemic effects in a dyslipidemia murine model. *Hum. Gene Ther.* **11**, 2415–2430 (2000).
13. Linton, M. F. *et al.* Transgenic mice expressing high plasma concentrations of human apolipoprotein B100 and lipoprotein(a). *J. Clin. Invest.* **92**, 3029–3037 (1993).
14. Purcell-Huynh, D. A. *et al.* Transgenic mice expressing high levels of human apolipoprotein B develop severe atherosclerotic lesions in response to a high-fat diet. *J. Clin. Invest.* **95**, 2246–2257 (1995).
15. Jackson, A. L. *et al.* Expression profiling reveals off-target gene regulation by RNAi. *Nature Biotechnol.* **21**, 635–637 (2003).
16. Bridge, A. J., Pebernard, S., Ducraux, A., Nicoulaz, A.-L. & Iggo, R. Induction of an interferon response by RNAi vectors in mammalian cells. *Nature Genet.* **34**, 263–264 (2003).
17. Llave, C., Xie, Z., Kasschau, K. D. & Carrington, J. C. Cleavage of Scarecrow-like mRNA targets directed by a class of Arabidopsis miRNA. *Science* **20**, 2053–2056 (2002).
18. Yekta, S., Shih, I. H. & Bartel, D. P. MicroRNA-directed cleavage of HOXB8 mRNA. *Science* **304**, 594–596 (2004).
19. Damha, M. J. & Ogilvie, K. K. Oligoribonucleotide synthesis. The silyl-phosphoramidite method. *Methods Mol. Biol.* **20**, 81–114 (1993).
20. Iyer, R. P., Egan, W., Regan, J. B. & Beaucage, S. L. 3H–1,2-benzodithiole-3-one 1,1-dioxide as an improved sulfurizing reagent in the solid phase synthesis of oligodeoxyribonucleoside phosphorothioates. *J. Am. Chem. Soc.* **112**, 1253–1254 (1990).
21. Zlot, C. H. *et al.* Generation of monoclonal antibodies specific for mouse apolipoprotein B-100 in

apolipoprotein B-48-only mice. *J. Lipid Res.* **40**, 76–84 (1999).

22. Hammad, S. M. *et al.* Lipoprotein subclass profiles of hyperlipidemic diabetic mice measured by nuclear magnetic resonance spectroscopy. *Metabolism* **52**, 916–921 (2003).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank P. Sharp for his advice and creative input. We are grateful to J. Maraganore and T. Ulich for their support and encouragement. We would like to thank S. Young for the LF3 anti-mouse apoB antibody; D. Bartel and S. Yekta for advice on the 5'-RACE assay; S. Young and M. Stoffel for valuable discussions; and LipoFIT Analytic GmbH and the Institute for Biophysics and Physical Biochemistry of the University of Regensburg for the characterization of lipoprotein particles by NMR. For technical assistance we thank P. Deuerling, F. Hertel, S. Leuschner, N. Linke, A. Müller, G. Ott, H. Schübel, S. Shanmugam, M. Duckman and C. Auger.

Competing interests statement The authors declare competing financial interests: details accompany the paper on Nature's website (<http://www.nature.com/nature>).

Correspondence and requests for materials should be addressed to J.S. (jsoutschek@alnylam.de) or H.-P.V. (hpvornlocher@alnylam.de).

The netrin receptor UNC5B mediates guidance events controlling morphogenesis of the vascular system

Xiaowei Lu^{1*}, Ferdinand le Noble^{2*}, Li Yuan^{2*}, Qingjan Jiang^{2†}, Benjamin de Lafarge², Daisuke Sugiyama³, Christiane Bréant², Filip Claes⁴, Frederik De Smet⁴, Jean-Léon Thomas⁵, Monica Autiero⁴, Peter Carmeliet⁴, Marc Tessier-Lavigne^{1†} & Anne Eichmann²

¹Department of Biological Sciences, Howard Hughes Medical Institute, Stanford University, Stanford, California 94305, USA

²INSERM U 36, Collège de France, 11 Place Marcelin Berthelot, 75005 Paris, France

³CNRS UMR 7622, Université Pierre et Marie Curie, 9 Quai St Bernard, 75252 Paris, France

⁴Center for Transgene Technology & Gene Therapy, Flanders Interuniversity Institute for Biotechnology (VIB), University of Leuven, Herestraat 49B-3000, Leuven, Belgium

⁵Inserm U495, IFR Neurosciences, Hôpital de la Salpêtrière, Université Pierre et Marie Curie, 75651 Paris, France

* These authors contributed equally to this work

† Present addresses: College of Animal Science, South China Agricultural University, Guangzhou 510642, China (Q.J.); Genentech Inc., 1 DNA Way, South San Francisco, California 94080, USA (M.T.-L.)

Blood vessels and nerves are complex, branched structures that share a high degree of anatomical similarity. Guidance of vessels and nerves has to be exquisitely regulated to ensure proper wiring of both systems. Several regulators of axon guidance have been identified and some of these are also expressed in endothelial cells; however, the extent to which their guidance functions are conserved in the vascular system is still incompletely understood. We show here that the repulsive netrin receptor UNC5B is expressed by endothelial tip cells of the vascular system. Disruption of the *Unc5b* gene in mice, or of *Unc5b* or *netrin-1a* in zebrafish, leads to aberrant extension of endothelial tip cell filopodia, excessive vessel branching and abnormal navigation. Netrin-1 causes endothelial filopodial retraction, but only when UNC5B is present. Thus, UNC5B functions as a repulsive netrin receptor in endothelial cells controlling morphogenesis of the vascular system.

During development, blood vessels navigate along stereotyped paths towards their targets—similar to axonal growth cones¹. However, the mechanisms regulating vessel navigation remain incompletely understood. Specialized endothelial tip cells at the leading front of such vessels extend numerous filopodia, and respond to chemoattractant and repellent guidance cues that act over short or long range^{2,3}. The existence of such endothelial ‘growth cones’ highlights the anatomical similarities between the nervous and vascular systems¹. Axon guidance cues include members of four prominent families: netrins, semaphorins, ephrins and slits⁴. Endothelial cells express some of these molecules or their receptors, including the semaphorin receptors neuropilin-1 (refs 5, 6), neuropilin-2 (ref. 7) and plexin-D1 (refs 8–10); ephrinB2 (refs 11, 12) and its receptor EphB4 (ref. 13); and Robo4, a member of a slit receptor family (ref. 14). At least some of these factors regulate vessel path-finding^{9,10}. However, netrins and their receptors have not yet been implicated in vascular development.

Netrins are a conserved family of laminin-related molecules^{15–17}. They are secreted by midline cells and attract axons towards the midline, while repelling a subset of axons migrating away from the midline. Netrins act through receptors in two distinct families: the deleted in colorectal cancer (DCC)¹⁸ and UNC-5 families^{19–24}. Attraction is mediated by DCC receptors²⁵, whereas repulsion requires receptors of the UNC-5 family, acting alone or together with DCC family receptors. Thus, in *Caenorhabditis elegans*, UNC-40/DCC is required for migration towards ventral midline cells expressing the netrin UNC-6, but it also participates together with UNC-5 in migration away from UNC-6 (ref. 19). In *Xenopus*, ectopic expression of UNC-5 family receptors converts DCC-mediated attraction of axons into repulsion²⁴. In *Drosophila*, ectopic expression of UNC-5 mediates short-range axon repulsion, whereas long-range repulsion also requires Frazzled/DCC²⁶. Four mammalian homologues of UNC-5 have been identified: UNC5A–D^{22,23,27}

(previously known as UNC5H1–4). Disruption of *Unc5c* in the rostral cerebellar malformation (*Rcm*) mutant mouse causes abnormalities in migration of cerebellar granule and Purkinje cells²³, consistent with a role in repulsive guidance²⁸. In *Xenopus*, UNC5A and UNC5B convert netrin attraction to repulsion²⁴, but the *in vivo* functions of these receptors and UNC5D in mammals have not been defined thus far.

Selective expression of *Unc5b* in the vascular system

By *in situ* hybridization in mouse embryos between embryonic day (E)10.5 and E18.5, *Dcc* was detectable in commissural neurons in the dorsal neural tube¹⁸ and other previously reported locations^{27,29–31}, but not in endothelial cells (Fig. 1a). *Unc5a* was detectable in various neural and non-neural structures^{22,27,32} (not shown), including the otic placode and olfactory system, but not in the vasculature. Importantly, *Unc5b*, although weakly expressed in the ventricular zone of the neural tube, strongly labelled capillaries in the perineural vascular plexus and those invading the neural parenchyma (Fig. 1b). *Unc5b* was expressed in the dorsal neural retina, but also strongly labelled the developing intra-ocular vasculature (Fig. 1d). Both *Dcc* (not shown) and *Unc5a* were expressed in the neural retina, but not in developing capillaries (Fig. 1c). Thus, *Unc5b* was expressed in developing blood vessels in all tissues at all stages, whereas neither of the other two netrin receptors was detected in the vascular system. Moreover, the expression of *Unc5b* appeared largely restricted to the vascular system, although expression in some other (non)-neural structures was also observed^{22,27,33}.

We next used *Unc5b* knockout mice, in which we engineered a β -galactosidase reporter into the *Unc5b* locus (see below), to visualize *Unc5b*-expressing cells. Whole-mount 5-bromo-4-chloro-3-indolyl- β -D-galactoside (X-gal) staining of E10.5 heterozygous embryos confirmed that *Unc5b* was mainly expressed in the vascular system, with neural expression restricted until E12.5 to the developing

retina and otic placode (Fig. 1e). Notably, expression was not uniform throughout the vasculature: the dorsal aorta, intersomitic arteries and internal carotid artery were positive, whereas superficially localized veins were negative (Fig. 1e, f). Double staining with X-gal and isolectinB4 confirmed that arteries, but not veins, expressed *Unc5b* in the retinal vasculature of postnatal *Unc5b* heterozygous pups (Fig. 1g–j; see also Supplementary Fig. 1). *Unc5b* expression in capillaries extended approximately halfway between the arteries and veins. High-magnification double staining of retinal vessels with isolectinB4 and an anti- β -galactosidase antibody showed that endothelial cells, but not pericytes, expressed *Unc5b* (Fig. 1i; see also Supplementary Fig. 1). Notably, endothelial tip cells at the forefront of arterial and venous sprouts strongly expressed *Unc5b* (Fig. 1i, j; see also Supplementary Fig. 1). Thus, *Unc5b* was expressed by arterial and a subset of capillary endothelial cells, as well as by endothelial tip cells.

More vessel branches and tip-cell filopodia in *Unc5b* mutants

Inactivation of the *Unc5b* gene was achieved by homologous recombination of a “secretory trap” vector³⁴ into the second intron of the *Unc5b* locus (Supplementary Fig. 2). This secretory trap

vector possesses a splice acceptor followed by a linker protein, a transmembrane domain, β -geo (a β -galactosidase/neomycin phosphotransferase fusion), an internal ribosome entry site, and placental alkaline phosphatase. As a result of the integration, upstream exons splice to the vector sequences, resulting in the generation of a fusion protein between part of the first immunoglobulin domain of UNC5B and the vector-coded sequences. No wild-type transcripts were detected in the mutants by northern blot analysis and—as detected by real-time polymerase chain reaction with reverse transcription (RT-PCR) analysis—*Unc5b* transcripts in the mutant mice were on average only $3 \pm 2\%$ ($\pm$ s.d.) of wild-type levels (Supplementary Fig. 2). As β -geo fusion proteins are retained in an intracellular compartment³⁴, this mutation is thus expected to generate a severely hypomorphic or null allele. All heterozygous animals were indistinguishable from wild type. On a C57BL/6 background, homozygous embryos were recovered at expected mendelian frequencies until E10, when they appeared growth-retarded and often presented cephalic neural tube closure defects (not shown). On an outbred CD1 background, however, growth retardation and failure of neural tube closure were not observed (not shown), and homozygotes were recovered at expected

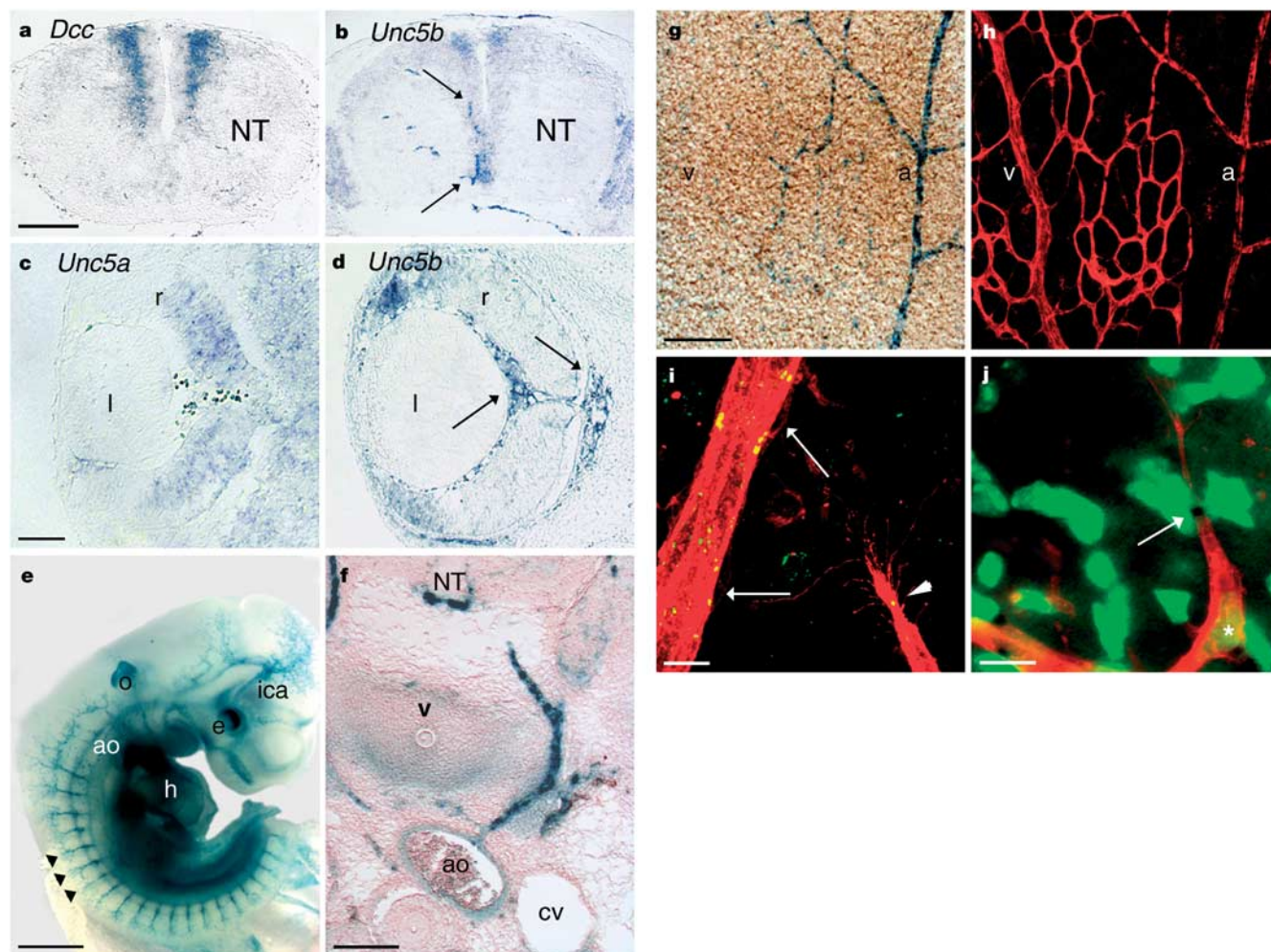


Figure 1 *Unc5b* expression in the vascular system. **a–d**, *In situ* hybridization at E12.5 of neural tube (**a**, **b**) and eye (**c**, **d**). Expression of *Dcc* in dorsal neural tube (NT; **a**), *Unc5a* in retina (**c**), and *Unc5b* in capillaries (**b**, **d**; arrows). **e**, X-gal staining at E10.5 of *Unc5b*^{+/-} mice. Note expression in dorsal aorta (ao), internal carotid artery (ica) and intersomitic arteries (arrowheads). Neural expression of *Unc5b* is restricted to the otic placode (o) and eye (e). **f**, Section of X-gal-stained E12.5 *Unc5b*^{+/-} embryo. Dorsal aorta and its branches are labelled; cardinal vein (cv) is negative for stain. **v**,

vertebra. **g**, **h**, LacZ (**g**) and isolectinB4 (**h**) double-labelling of retinal vasculature in P4 *Unc5b*^{+/-} mice. Arteries (a) but not veins (v) express *Unc5b*. **i**, isolectinB4 (red) and anti- β -galactosidase antibody (green) double-labelling. Weakly positive isolectinB4 pericytes lining the arterial stem (arrows) are *Unc5b*-negative. Arrowhead indicates a β -galactosidase-positive endothelial tip cell. **j**, X-gal-positive tip cell (arrow). The nucleus (asterisk) is stained with Hoechst (green). Scale bars: **a**, **c**, **f**, **g**, 0.1 mm; **e**, 1 mm; **i**, **j**, 10 μ m.

mendelian frequencies until E12.5. Their death was due to heart failure, as indicated by accumulation of blood in the venous circulation and fluid in the pericardial cavity. As no obvious cardiac malformations were detected (not shown), the heart failure probably resulted from the increased peripheral resistance due to the abnormal arterial vasculature (described below), although alternative possibilities cannot be excluded. The vascular phenotype of C57BL/6 and CD1 embryos before E10 was very similar (not shown).

At E8–E10, homozygous *Unc5b* mutant embryos initially formed a normal vascular plexus, which remodelled into arteries and veins

(Figs 2 and 3). However, whole-mount X-gal staining revealed increased vessel branching of the internal carotid artery, inter-somitic vessels and vessels in the nervous system in mutant embryos (Fig. 2a–f). Double labelling of brain capillaries by whole-mount staining with X-gal and isolectinB4 (Fig. 2g, h) confirmed that capillaries of mutant hindbrains appeared thinner and more highly branched than vessels in stage-matched wild-type or heterozygous embryos (Fig. 2i–k). Vessels in the hindbrain (Fig. 2l) and neural tube (not shown) had 40% more branch points in homozygous than heterozygous mutants, when counted after isolectinB4 staining. Interestingly, *Unc5b*-expressing tip cells had more filopodial

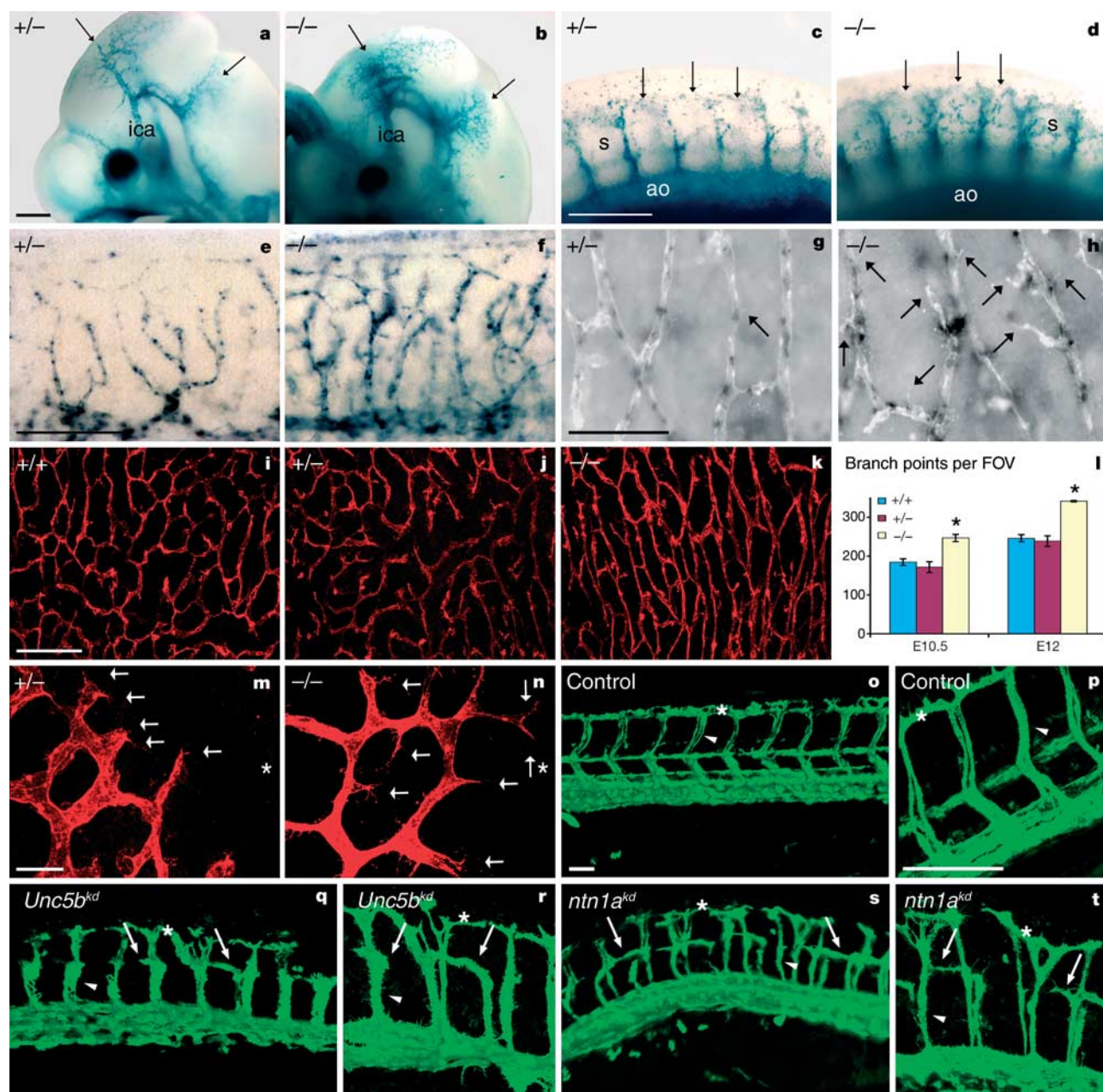


Figure 2 Increased capillary branching and tip-cell filopodia extension in *Unc5b* mouse and zebrafish mutant embryos. **a–f**, Whole-mount X-gal staining of E10.5 *Unc5b* ^{+/-} and *Unc5b* ^{-/-} littermates. Homozygous embryos show increased vessel branching from the internal carotid artery (ica) (**b**; arrows), in the somitic region (s) (**d**; arrows) and in the neural tube (**f**). ao, dorsal aorta. **g, h**, LacZ and isolectinB4 double-staining of virtually all endothelial cells in E11 neural tube. Arrows indicate increased tip-cell filopodial extensions in *Unc5b* ^{-/-} capillaries (**h**). **i–k**, isolectinB4 staining of hindbrain vessels in E12 mice. **l**, Quantification of branch point number per field of view (FOV). Bars represent

mean $\pm$ s.e.m., asterisk, $P < 0.05$ in this and all subsequent figures (Mann–Whitney *U*-test). **m, n**, High-magnification of E10.5 tip cells at the hindbrain midline (asterisk). Filopodial extensions (arrows) are increased in **n**. **o–t**, Lateral views of zebrafish embryos, head on the left. In control morphants (**o, p**), ISVs (arrowheads) form the DLAV (asterisk). In *Unc5b* (**q, r**) and *ntn1a* (**s, t**) morphants, vessel branching is increased (**q–t**; arrows). Note increased filopodial extension (**r, t**). Scale bars: **a–d**, 0.5 mm; **e, f**, 0.2 mm; **g–k**, 0.1 mm; **m, n**, 20 μ m; **o, q, s**, 50 μ m; **p, r, t**, 75 μ m.

extensions in homozygous than heterozygous mutant embryos (Fig. 2g, h). High-magnification confocal microscopy confirmed that filopodial extension from tip cells was markedly increased in *Unc5b* mutant embryos (Fig. 2m, n).

Aberrant vessels in *Unc5b* and *netrin-1a* morphant zebrafish

We next examined whether perturbing *Unc5b* in zebrafish embryos affects development of intersegmental blood vessels (ISVs), as their path-finding is stereotyped and directed by guidance cues^{10,35}. To visualize ISVs, we used *Tg(fli1:EGFP)^{y1}* zebrafish, which expressed enhanced GFP in their endothelial cells³⁶. In embryos injected with buffer or control antisense morpholino oligonucleotides (morpholinos) (Fig. 2o, p), primary ISVs sprouted from the dorsal aorta around 20 h post-fertilization (h.p.f.) and extended dorsally to the roof of the neural tube, with the dorsally positioned endothelial tip cell extending rostral and caudal filopodia that subsequently fused with neighbouring cell extensions to establish, by 48 h.p.f., the

dorsal longitudinal anastomosing vessel (DLAV)³⁶. To knockdown the zebrafish *Unc5b* orthologue, we used a splice-site-specific morpholino, which lowered the levels of normally spliced *Unc5b* messenger RNA in morphant *Unc5b^{kd}* embryos (Supplementary Fig. 3). As netrin is a ligand for UNC5B, we also knocked down the zebrafish *netrin-1a* (*ntn1a*) orthologue³⁷. The knockdown of *ntn1a* (*ntn1a^{kd}*) markedly phenocopied the *Unc5b* knockdown.

Unc5b^{kd} and *ntn1a^{kd}* embryos had normal somite numbers by 24 h.p.f. Formation and perfusion of the axial trunk vessels (that is, the dorsal aorta and posterior cardinal vein) proceeded normally, indicating that arterio-venous specification was not affected. Initial sprouting of the primary ISVs from the dorsal aorta into the intersegmental space occurred normally in *Unc5b* and *ntn1a* morphants, although with a delay of 6 h. Notably, however, by 48 h.p.f. knockdown of *Unc5b* caused dose-dependent vessel-branching defects, affecting 30% or 70% of morphant embryos after injection of 2 or 6 ng of *Unc5b* morpholino per embryo, respectively (Fig. 2q, r). Corresponding numbers for the *ntn1a* knockdown were 48% or 71% of morphant embryos at 2 or 3 ng of *ntn1a* morpholino per embryo (Fig. 2s, t). In the most affected embryos, up to 70% of the primary ISVs showed abnormal trajectory phenotypes (Fig. 2q–t). ISV endothelial cells exhibited supernumerary and ectopic filopodial extensions alongside the entire vessel path in *Unc5b^{kd}* and *ntn1a^{kd}* embryos, whereas in control embryos, the leading endothelial tip cell showed the most extensive filopodial extensions (Fig. 2r, t). The dorsal trajectory of most ISVs was irregular and deviated from the normal stereotyped path found in control embryos. In mild cases, the ISVs were normally positioned and only showed supernumerary filopodial extensions. In more affected ISVs, these filopodia formed an extra vessel branch, emerging horizontally at variable positions alongside the ISV, projecting in a rostral or caudal direction and ectopically fusing with a neighbouring ISV to form an aberrant anastomosis (Fig. 2q–t). Primary ISVs are known to form by migration—not proliferation—of endothelial cells, budding off dorsally from the dorsal aorta³⁶. Counting the number of endothelial nuclei in the *Unc5b^{kd}* and *ntn1a^{kd}* fish failed to reveal a significant increase in cell number per ISV (not shown), suggesting that the observed phenotype was due to aberrant guidance rather than altered cell proliferation. Overall, zebrafish *Unc5b* and *ntn1a* morphants phenocopied the vascular defects observed in homozygous *Unc5b* mouse embryos.

Selective vessel branching defects in *Unc5b* mutant mice

To confirm whether, as in zebrafish, the vessel phenotype is due to endothelial branching and navigation but not proliferation defects, *Unc5b* mouse mutant embryos were stained with different markers. Staining with the pan-vascular marker PECAM-1 showed that the lumen of abnormally branched mutant arteries was often collapsed or irregularly shaped (Fig. 3a, b). The affected arteries never included the dorsal aorta, but did include its side branches (Fig. 3). In spite of their abnormal morphology, homozygous mutant vessels showed normal expression of arterial markers ephrinB2 (not shown) and neuropilin-1 (Fig. 3c, d) and of the venous markers neuropilin-2 and *EphB4* (not shown). The vessel wall of mutant arteries also formed normally, as indicated by expression of *Pdgfrb* (Fig. 3e, f) or anti-smooth-muscle actin staining (not shown). Endothelial proliferation, judged by counting of cells doubly positive for BrdU-LacZ, was not significantly different between heterozygous and homozygous embryos (Fig. 3g–k). Apoptosis of endothelial cells was virtually undetectable in embryos of either genotype (Fig. 3l; and not shown). Taken together, these observations suggest that the primary defect in *Unc5b* mutant vessels is a branching defect.

Netrin-1 reduces endothelial migration and filopodial extension

Because *Unc5b*-deficient vessels in mice and zebrafish exhibit

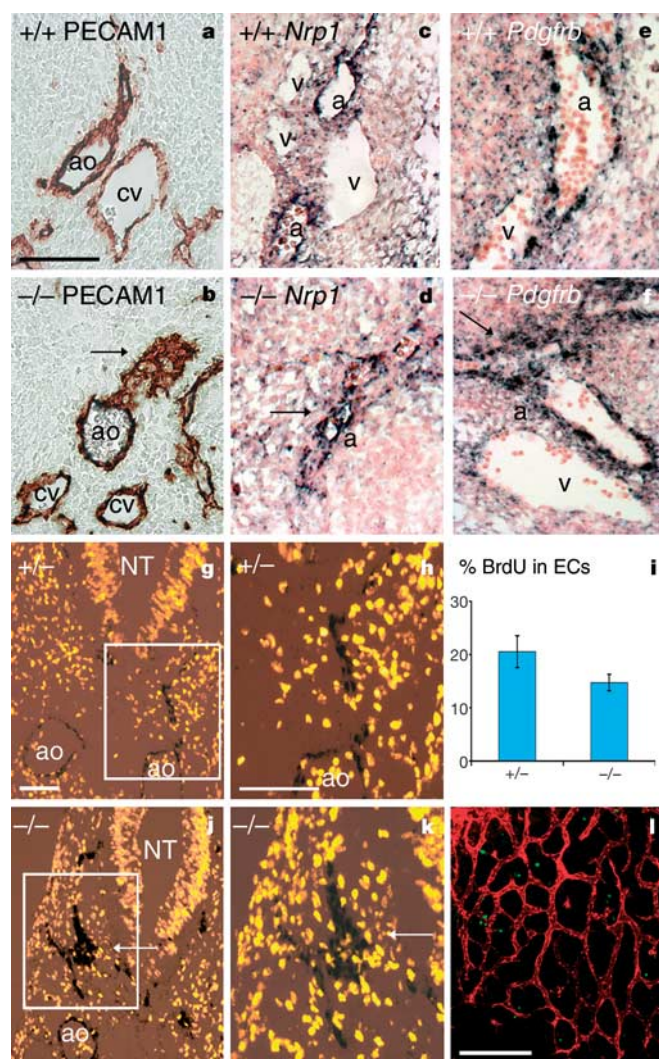


Figure 3 Abnormal morphology of *Unc5b* mouse mutant arteries. **a, b**, Tail region at E11. **b**, Collapsed lumen of aorta (ao) branch (arrow). The cardinal vein (cv) appears normal. **c–f**, *In situ* hybridization. In spite of abnormal lumen (**d, f**; arrows), *Unc5b^{-/-}* arteries (a) express neuropilin-1 (*Nrp1*) (**d**) and *Pdgfrb* (**f**) normally. v, vein. **g–k**, Normal proliferation of *Unc5b* mutant vessels. **g, h, j, k**, BrdU (yellow) and LacZ (black) double staining. **h, k**, Higher magnification of boxed regions in **g, j**. Note abnormal lumen in *Unc5b^{-/-}* intersomitic artery (**j, k**; arrows). **i**, Quantification of BrdU-LacZ double-labelled endothelial cells (ECs). **l**, Cleaved caspase-3 (green) and isolectinB4 (red) double staining of hindbrain vessels. There is no endothelial cell apoptosis. Scale bar, 0.1 mm.

ectopic filopodial extensions, UNC5B seems to negatively regulate filopodial extension in the vascular system, which could result in inhibition of cell migration; we thus examined the activity of the UNC5B ligand netrin-1 on endothelial cell migration, using transwell and wound migration assays. Using RT-PCR, we found that *UNC5B* was strongly expressed in primary human umbilical artery endothelial cells (HUAECs) and, at lower levels, in human umbilical vein endothelial cells (HUVECs); *DCC*, *UNC5A*, *-C* and *-D* were undetectable and mRNA for the adenosine 2b receptor, which also binds netrin-1 (ref. 38), was expressed at equal levels (Fig. 4a, b; see also Supplementary Fig. 3). Medium conditioned by stably transfected 293 cells secreting netrin-1 (ref. 39) (containing $\sim 2.5 \mu\text{g ml}^{-1}$ netrin-1; not shown) did not affect transwell migration of HUVECs but reduced migration of HUAECs by $\sim 40\%$ (Fig. 4c). Migration of 'wounded' confluent HUVECs was also dose-dependently inhibited by netrin-1 (Fig. 4d). Thus, netrin-1 inhibited migration of endothelial cells expressing *UNC5B* *in vitro*, consistent with a possible negative role in filopodial extension.

To document the effects of netrin-1 on filopodial extension, we

used aortic ring sprouting assays (Fig. 4e–h; see also Supplementary Fig. 4). Time-lapse video microscopy of control sprouting endothelial tip cells over 2 h showed little or no net forward or reverse movement of the tip (Fig. 4e, f; see also Supplementary Fig. 4 and Supplementary Movie 1). In contrast, exposure of endothelial tip cells to a gradient of netrin-1 resulted in clear retraction of the tip cell filopodia and backward movement of the tip cell (Fig. 4g, h; see also Supplementary Fig. 4 and Supplementary Movie 3). Pre-incubation of netrin-1 with recombinant UNC5B-Fc, a fusion of the constant region of human IgG with the ectodomain of UNC5B (known to bind netrin-1 (ref. 22)), blocked filopodial retraction and backward movement of the tip cell (Supplementary Fig. 4 and Supplementary Movie 2).

To examine whether netrin-1 also affected filopodial extension *in vivo*, we performed intra-ocular injections of recombinant protein into postnatal day 5 (P5) mice, followed by analysis of the retinal vasculature. Neither netrin-1 (*Ntn1*) nor netrin-4 (*Ntn4*) expression could be detected in P5 retinal whole mounts by *in situ* hybridization (not shown), although we observed embryonic *Ntn1*

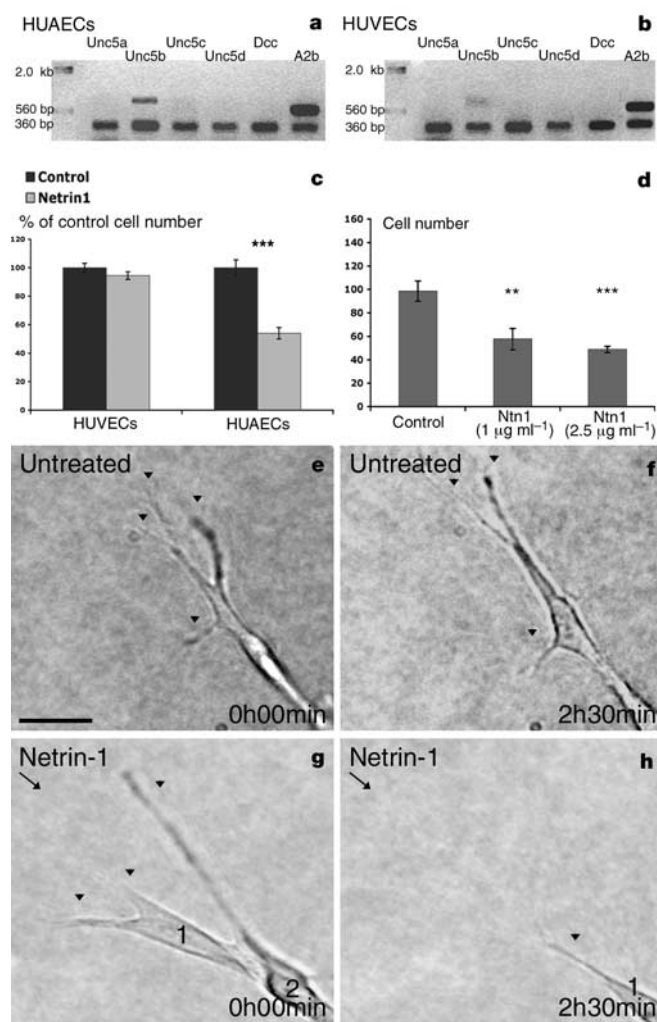


Figure 4 Netrin-1 reduces endothelial cell migration and filopodial extension *in vitro*. **a, b**, Expression of netrin receptors and β -tubulin (bottom band) in HUAECs (**a**) or HUVECs (**b**). **c**, Transwell migration assay. **d**, HUAEC wound migration assay. **e, f**, Representative examples of three experiments. **e–h**, Aortic ring sprouting assay. Still images from time-lapse videos taken at the indicated time points. Arrowheads indicate endothelial tip cell filopodia. Note little net movement of untreated tip cells (**e, f**). **g, h**, Exposure of two tip cells (1, 2) to a netrin-1 gradient (source indicated by arrows) induces rapid filopodial retraction and backward movement of both cells. Scale bar, 30 μm .

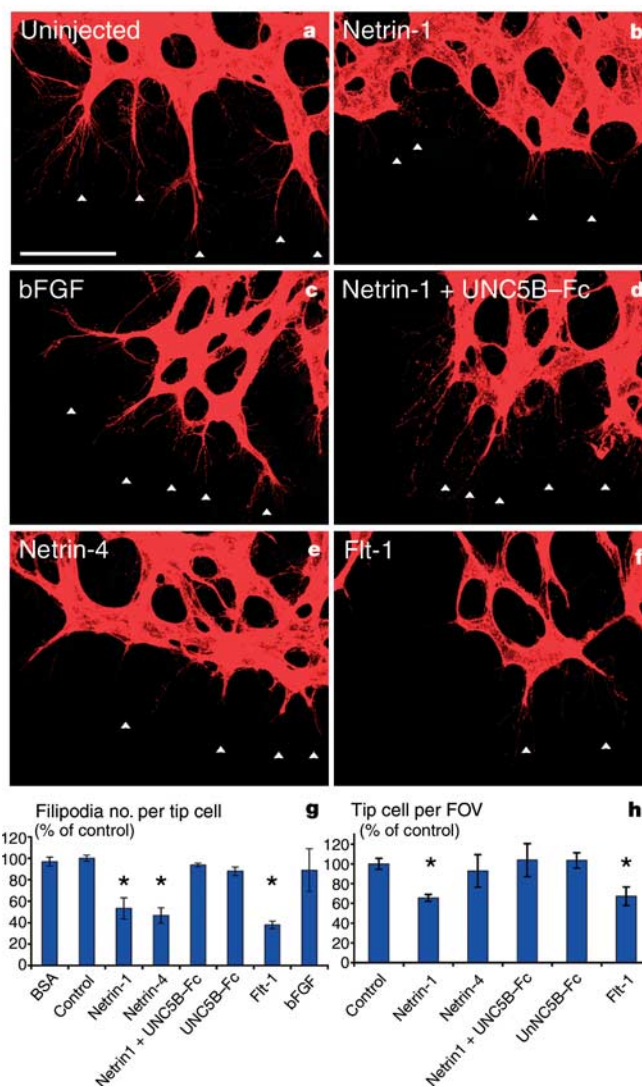


Figure 5 Netrin-1 stimulates filopodial retraction of endothelial cells *in vivo*. **a–f**, Confocal images of isolectinB4-stained tip cells (arrowheads) at the angiogenic front in control eyes (**a**) or eyes injected with the indicated proteins (**b–f**). Note filopodial retraction in **b, e, f**. Scale bar, 50 μm . **g**, Quantification of filopodia number per tip cell. **h**, Quantification of tip cell number per FOV.

expression in the optic nerve head as previously described³⁰. Compared to uninjected control eyes, netrin-1-injected eyes showed a marked decrease in filopodial extension over the entire angiogenic front (Fig. 5a, b). Both the number of filopodia-extending cells and filopodia per tip cell decreased significantly (Fig. 5g, h), such that the netrin-1-injected angiogenic front appeared smooth compared to the 'spiky' control front (Fig. 5a, b). Filopodial retraction induced by netrin-1 was specific, as filopodia were not affected by control BSA (Fig. 5g) or the growth factor bFGF, a protein that, like netrin-1, is highly basic and binds heparin tightly (Fig. 5c, g). Filopodial retraction induced by netrin-1 was completely neutralized by pre-incubation of netrin-1 protein with recombinant UNC5B-Fc (Fig. 5d, g, h). Injection of recombinant UNC5B-Fc protein alone had no effect on filopodia, consistent with the absence of expression of endogenous netrin at this developmental stage (Fig. 5g, h). Recombinant netrin-4 (ref. 40) also induced filopodial retraction, although the effect was weaker (Fig. 5e, g, h). Acute sequestration of

VEGF by injection of soluble Flt-1 (ref. 2) (Fig. 5f–h) was as effective at stimulating filopodial retraction as was injection of netrin-1.

UNC5B is required for netrin-induced filopodial retraction

To assess whether filopodial retraction induced by netrin-1 is mediated by signalling through UNC5B, we injected recombinant proteins into hindbrains of E10.5 *Unc5b* mutant embryos, followed by 3-h embryo culture and analysis of the vasculature. Tip cell morphology was analysed at the dorsal hindbrain, furthest removed from the floor plate, where endogenous netrin-1 is produced¹⁵. Injection of netrin-1 in wild-type or heterozygous embryos resulted in a marked reduction of filopodial extension from tip cells compared with uninjected or BSA-injected controls (Fig. 6a, b). As in the eye, the dorsal angiogenic front of netrin-1-injected capillaries appeared smooth, in contrast to its spiky appearance in uninjected or BSA-injected embryos. At high magnification, residual tip cells in wild-type or heterozygous netrin-1-injected embryos showed few remaining filopodia (Fig. 6h); the number of tip cells extending filopodia in those animals was decreased by 30% compared with uninjected controls (Fig. 6e).

The dorsal angiogenic front in uninjected *Unc5b* mutant hindbrains showed more tip cells than wild-type or heterozygous controls, reflecting the increased capillary branching in these embryos (Fig. 6c, e). At high magnification, filopodial extension from dorsal tip cells was increased in uninjected mutant embryos compared with controls (Fig. 6f, g) but, strikingly, netrin-1 injection did not change capillary morphology (Fig. 6c–e), the number of filopodia-extending tip cells (Fig. 6d, e), or filopodial extension of mutant tip cells (Fig. 6i, j). Taken together, these observations suggest that netrin-1-induced filopodial retraction of endothelial tip cells is mediated by UNC5B signalling.

Discussion

To produce the highly stereotyped vascular pattern observed in adults, branching of the developing vascular system, initially laid down as a honeycomb-like plexus of uniformly sized small vessels, must be intricately regulated. Most importantly, extension of capillary tip cell filopodia requires sequestration of the pro-angiogenic factor VEGF to the extracellular matrix^{2,3}. We have provided evidence that netrin-1, acting via its receptor UNC5B, is a negative regulator of capillary branching in the developing vascular system. Loss of function of *Unc5b* in mice leads to ectopic vessel branching in the brain, head and somitic region and to malformation of numerous intra-embryonic arteries. Knockdown of *Unc5b* and *ntn1a* in zebrafish also leads to ectopic branching and path-finding defects of ISVs. UNC5B has also been described as a dependence receptor, mediating apoptosis in the absence of ligand^{41,42}. If so, loss of UNC5B might be expected to decrease apoptosis, and loss of its ligand to increase it. However, apoptosis of endothelial cells in the early developing vascular system is extremely rare⁴³, and no obvious changes in proliferation, endothelial marker expression or vessel wall assembly were detected in *Unc5b* mutant mice, or in *Unc5b* or *ntn1a* morphants in zebrafish, suggesting that *Unc5b* selectively regulates endothelial guidance, not survival. Loss of *Unc5b* function in capillary growth cones leads to excessive filopodial extension and increased branching, suggesting that its normal function is to negatively regulate these guidance processes. It thus seems that during evolution, UNC5B has been co-opted by the vascular system to function as a repulsive guidance receptor for vascular patterning.

In zebrafish, knockdown of *ntn1a* phenocopies the ISV guidance defects observed in *Unc5b*-deficient vessels. The vascular phenotype obtained after *ntn1a* disruption in zebrafish is consistent with the absence of repellent guidance function in the vasculature, as loss of *ntn1a* in the somites leads to ectopic tip cell extension at the level of the horizontal myoseptum. Our data also support the idea that in mice UNC5B functions as a receptor for netrin-1 (ref. 22) *in vivo*: injections of recombinant protein into the hindbrain induced

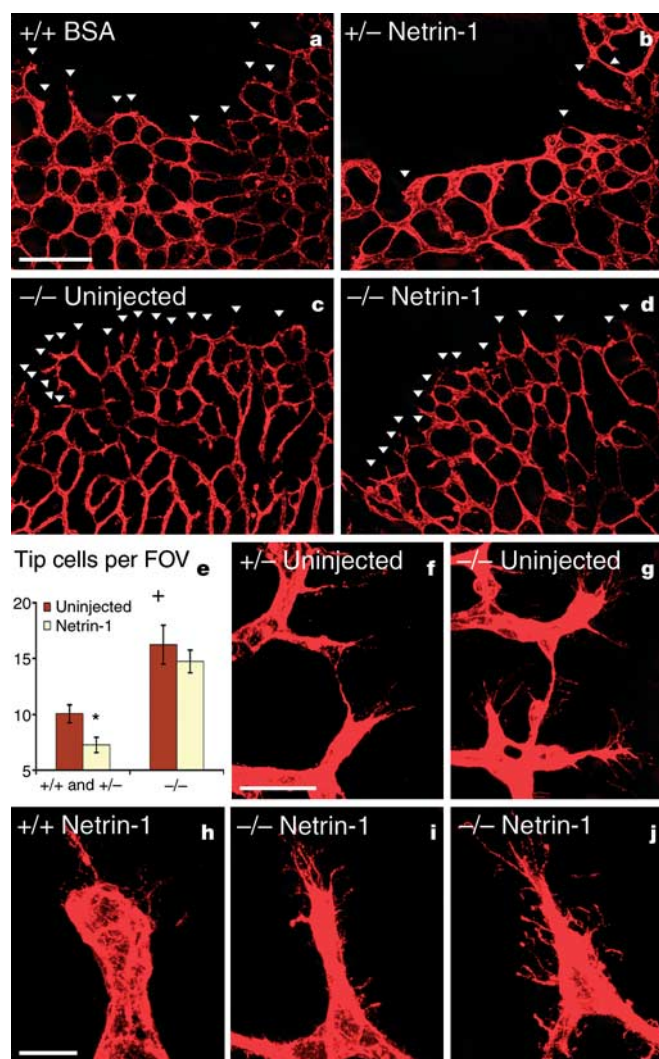


Figure 6 Netrin-1-induced filopodial retraction is lost in the absence of *Unc5b*. Confocal images of isolectinB4-stained capillaries of E10.5 hindbrains injected with recombinant proteins. **a–d**, Dorsal angiogenic front. Arrowheads indicate tip cells extending filopodia. **e**, Number of tip cells per FOV; note the 30% reduction in netrin-1-injected wild-type and heterozygous embryos (asterisk, $P < 0.05$). Uninjected *Unc5b*^{-/-} embryos show increased tip cell number compared with uninjected wild-type or heterozygous embryos (plus symbol, $P < 0.05$), but no significant reduction in tip cell number after netrin-1 injection. **f–j**, Higher magnifications of filopodial extension from tip cells (injected or not as indicated). Scale bars: **a–d**, 100 μ m; **f, g**, 20 μ m; **h–j**, 10 μ m.

filopodial retraction in wild type but not in *Unc5b* mutant endothelial growth cones. Netrin-1 is expressed in the floor plate and somites of mouse embryos¹⁵, consistent with a function as an endogenous UNC5B ligand. Netrin-1 has a high affinity for cell membranes^{15,16}, and is thought to be diffusible in some environments (such as the embryonic spinal cord¹⁵) but not others (like the optic nerve head³⁰). In the case of the vasculature, it is not clear whether the ligand that is presumed to stimulate UNC5B on endothelial cells acts over a short or long range. Examination of *Ntn1* mutant mice has not, so far, revealed any obvious vascular malformations (not shown). One possibility is that additional netrins might act in a redundant manner and compensate for the lack of netrin-1 in mouse embryos. *Ntn4* is expressed in the lateral floor plate and dorsal root ganglia (DRG)⁴⁴, and DRG also express *Ntn3* (ref. 45). Netrin-4 can cause filopodial retraction after intracocular injections. Thus, different netrins may collaborate to regulate vascular development. Generation of double or triple knockout mice will determine which netrins function as endogenous UNC5B ligands in mice.

UNC5B is not only expressed by endothelial growth cones, but also by arterial endothelial cells. Once established, the arterial tree does not generate new branches during normal development, suggesting that UNC5B may have an ongoing physiological role in repressing branching in the adult as well. UNC5B functions as a negative regulator of capillary branching, controlling vascular morphogenesis. Examination of netrin/UNC5B expression and function during vascular pathologies may suggest roles as novel targets in pro- or anti-angiogenic therapies. □

Methods

Gene targeting at the *Unc5b* locus

A targeting vector was constructed by flanking the 'secretory gene trap' cassette³⁴ with 4.5 kilobases (kb) and 3.5 kb of *Unc5b* genomic fragments, amplified from E14 embryonic stem (ES) cells using the Expand High Fidelity PCR system (Roche). Details of vector cloning are available upon request. The targeting vector was electroporated into E14 ES cells. Homologous recombination events, detected by a 1.3-kb external probe, resulted in the insertion of the secretory gene trap cassette and the deletion of a 2-kb genomic region containing exons 3 and 4. From 54 ES cell clones screened by Southern blot, 11 clones were identified.

RNA analysis

Total RNA was isolated from individual E10.5 embryos using the RNeasy Mini kit (Qiagen) following the manufacturer's instructions. Real-time RT-PCR analysis was performed with SYBR green chemistry on a Stratagene Mx3000P system. Three complementary DNA samples per genotype were synthesized with oligo dT primers and Superscript II Reverse Transcriptase (Invitrogen). PCR reactions were carried out with cDNA synthesized from 17 ng of total RNA, 7.5 pmol of each primer, and 1 × SYBR green reaction mix (Applied Biosystems) in a 25 µl volume. *Unc5b* transcript levels in individual samples were first normalized to the level of the transferrin receptor, then plotted as a percentage of wild-type levels (100% is the mean of the wild-type levels).

Recombinant proteins and immunohistochemistry

All recombinant proteins were from R&D, except bFGF (Bachem). Histology, immunohistochemistry, *in situ* hybridization and X-gal staining were as described^{2,7}, using isolectinB4 (Sigma), streptavidin Cy-3 (Amersham), PECAM-1 (Pharmingen), anti-NRP-1 and anti-NRP-2 (R&D) and cleaved caspase-3 (Cell Signaling). BrdU injections were performed as described⁷; embryos were collected after 3 h. Confocal images were acquired using a Leica TCS SP2 confocal microscope. Two independent observers counted endothelial branch points. For each stage, three 1 × 1 mm images taken from four *Unc5b*^{-/-}, four *Unc5b*^{+/-} and three *Unc5b*^{+/+} embryos (four litters) were counted. For quantification of BrdU-LacZ double-labelled endothelial cells, six sections from four *Unc5b*^{-/-}, four *Unc5b*^{+/-} and two *Unc5b*^{+/+} embryos were counted using Metaview software (Princeton, version 5.0r6, 2002). All statistical analysis was done using the Mann-Whitney *U*-test; *P* < 0.05.

Morpholino knockdowns

Tg(fli1:EGFP)^{y1} zebrafish³⁶ were provided by Zebrafish International Resource Center (University of Oregon) and maintained as described⁴⁶. Morpholinos (Gene Tools; see Supplementary Fig. 3a) were injected into single- to four-cell-stage zebrafish embryos as described⁴⁷. About 40 embryos were analysed per experiment, and each experiment was repeated four–five times. The penetrance of the phenotype was scored by counting the affected embryos. Negative controls, including a morpholino against the zebrafish *six6* gene (*zsisix6*, 5'-CTCTAAAGGAGACCTGAAACCATG-3') did not produce a vascular phenotype. Confocal imaging was performed using a Zeiss laser-scanning microscope LSM510.

Endothelial migration assays

HUAECs and HUVECs (Promocell, passage 3 to 7) were cultured in endothelial cell growth medium (ECGM, Promocell) containing 10% growth supplement (Promocell). The upper wells of transwell migration chambers (Costar, 8-µm-pore filters, 50 µg ml⁻¹ fibronectin coating) were seeded with 5 × 10⁴ HUAECs or HUVECs. The lower chamber was seeded with confluent 293 or 293 netrin-1-secreting cells³⁹. Cells were left to migrate for 2 h; cells remaining in the upper well were mechanically removed. Cells at the bottom side of the filter were fixed with 4% paraformaldehyde, washed and counted after Hoechst nuclear stain using Metaview software (three experiments, duplicate wells). For wound-healing assays, confluent HUAECs starved overnight in 1% growth supplement were wounded with a pipette tip. After 24 h, cultures were photographed and two independent observers counted cells migrating into the wound area (three experiments, duplicate wells). For aortic ring assays, the abdominal aorta from 2-month-old anaesthetized rats was excised, washed in ECGM and cut into small rings. Rings were placed in semi-solid collagen cultures (Roche) prepared according to the manufacturer's instructions in ECGM containing 50 ng ml⁻¹ of VEGF for 5 days. Recombinant netrin-1 gradients (1 µg µl⁻¹) were applied using a micropipette (six experiments). Pre-clustering of netrin-1 with UNC5B-Fc was done at a 1:1 ratio for 1 h at 4 °C (three experiments). Time-lapse images of individual sprouts were acquired on an inverted microscope (Leica) equipped with a digital camera (Princeton Coolsnap cf) using Metaview software. Cell movement was measured by comparing the *x-y* positions of cell bodies at the beginning and after 2 h 30 min of time-lapse imaging.

Intraocular and hindbrain injections

Intraocular injections were performed as described⁷, except that pups were anaesthetized using ketamine-xylazine and killed 3 h after injection. All proteins were injected at 1 µg µl⁻¹ except bFGF (200 ng µl⁻¹). Number of injected eyes: netrin-1, five; netrin-1/UNC5B-Fc, three; netrin-4, three; Flt-1-Fc, two; bFGF, two; BSA, three; uninjected, 14. For hindbrain injections, E10.5 CD1 embryos were isolated in warm embryo culture medium⁴⁸, injected using calibrated micropipettes and cultured for 3 h in rat serum (Charles River) in a roller culture chamber (BTC Engineering). Number of injected embryos: BSA, four *Unc5b*^{+/+}, three *Unc5b*^{+/-}; netrin-1, three *Unc5b*^{+/+}, three *Unc5b*^{+/-}, four *Unc5b*^{-/-} (three litters). Quantification was done on three 1 mm/1 mm images per retina or hindbrain.

Received 12 August; accepted 6 October 2004; doi:10.1038/nature03080.

Published online 27 October 2004.

- Carmeliet, P. Blood vessels and nerves: common signals, pathways and diseases. *Nature Rev. Genet.* **4**, 710–720 (2003).
- Gerhardt, H. et al. VEGF guides angiogenic sprouting utilizing endothelial tip cell filopodia. *J. Cell Biol.* **161**, 1163–1177 (2003).
- Ruhrberg, C. et al. Spatially restricted patterning cues provided by heparin-binding VEGF-A control blood vessel branching morphogenesis. *Genes Dev.* **16**, 2684–2698 (2002).
- Dickson, B. J. Molecular mechanisms of axon guidance. *Science* **298**, 1959–1964 (2002).
- Kawasaki, T. et al. A requirement for neuropilin-1 in embryonic vessel formation. *Development* **126**, 4895–4902 (1999).
- Moyon, D., Pardanau, L., Yuan, L., Breant, C. & Eichmann, A. Plasticity of endothelial cells during arterial-venous differentiation in the avian embryo. *Development* **128**, 3359–3370 (2001).
- Yuan, L. et al. Abnormal lymphatic vessel development in neuropilin 2 mutant mice. *Development* **129**, 4797–4806 (2002).
- van der Zwaag, B. et al. PLEXIN-D1, a novel plexin family member, is expressed in vascular endothelium and the central nervous system during mouse embryogenesis. *Dev. Dyn.* **225**, 336–343 (2002).
- Gitler, A. D., Lu, M. M. & Epstein, J. A. PlexinD1 and semaphorin signaling are required in endothelial cells for cardiovascular development. *Dev. Cell* **7**, 107–116 (2004).
- Torres-Vazquez, J. et al. Semaphorin-plexin signaling guides patterning of the developing vasculature. *Dev. Cell* **7**, 117–123 (2004).
- Wang, H. U., Chen, Z. F. & Anderson, D. J. Molecular distinction and angiogenic interaction between embryonic arteries and veins revealed by ephrin-B2 and its receptor Eph-B4. *Cell* **93**, 741–753 (1998).
- Adams, R. H. et al. Roles of ephrinB ligands and EphB receptors in cardiovascular development: demarcation of arterial/venous domains, vascular morphogenesis, and sprouting angiogenesis. *Genes Dev.* **13**, 295–306 (1999).
- Gerety, S. S., Wang, H. U., Chen, Z. F. & Anderson, D. J. Symmetrical mutant phenotypes of the receptor EphB4 and its specific transmembrane ligand ephrin-B2 in cardiovascular development. *Mol. Cell* **4**, 403–414 (1999).
- Park, K. W. et al. Robo4 is a vascular-specific receptor that inhibits endothelial migration. *Dev. Biol.* **261**, 251–267 (2003).
- Kennedy, T. E., Serafini, T., de la Torre, J. R. & Tessier-Lavigne, M. Netrins are diffusible chemotropic factors for commissural axons in the embryonic spinal cord. *Cell* **78**, 425–435 (1994).
- Serafini, T. et al. The netrins define a family of axon outgrowth-promoting proteins homologous to *C. elegans* UNC-6. *Cell* **78**, 409–424 (1994).
- Cohen-Cory, S. The double life of netrin. *Nature Neurosci.* **5**, 926–928 (2002).
- Keino-Masu, K. et al. Deleted in Colorectal Cancer (DCC) encodes a netrin receptor. *Cell* **87**, 175–185 (1996).
- Hedgecock, E. M., Culotti, J. G. & Hall, D. H. The unc-5, unc-6, and unc-40 genes guide circumferential migrations of pioneer axons and mesodermal cells on the epidermis in *C. elegans*. *Neuron* **4**, 61–85 (1990).
- Leung-Hagsteejin, C. et al. UNC-5, a transmembrane protein with immunoglobulin and thrombospondin type 1 domains, guides cell and pioneer axon migrations in *C. elegans*. *Cell* **71**, 289–299 (1992).
- Hamelin, M., Zhou, Y., Su, M. W., Scott, I. M. & Culotti, J. G. Expression of the UNC-5 guidance receptor in the touch neurons of *C. elegans* steers their axons dorsally. *Nature* **364**, 327–330 (1993).
- Leonardo, E. D. et al. Vertebrate homologues of *C. elegans* UNC-5 are candidate netrin receptors. *Nature* **386**, 833–838 (1997).

23. Ackerman, S. L. *et al.* The mouse rostral cerebellar malformation gene encodes an UNC-5-like protein. *Nature* **386**, 838–842 (1997).
24. Hong, K. *et al.* A ligand-gated association between cytoplasmic domains of UNC5 and DCC family receptors converts netrin-induced growth cone attraction to repulsion. *Cell* **97**, 927–941 (1999).
25. Fazeli, A. *et al.* Phenotype of mice lacking functional Deleted in colorectal cancer (Dcc) gene. *Nature* **386**, 796–804 (1997).
26. Keleman, K. & Dickson, B. J. Short- and long-range repulsion by the *Drosophila* Unc5 netrin receptor. *Neuron* **32**, 605–617 (2001).
27. Engelkamp, D. Cloning of three mouse Unc5 genes and their expression patterns at mid-gestation. *Mech. Dev.* **118**, 191–197 (2002).
28. Przyborski, S. A., Knowles, B. B. & Ackerman, S. L. Embryonic phenotype of Unc5h3 mutant mice suggests chemorepulsion during the formation of the rostral cerebellar boundary. *Development* **125**, 41–50 (1998).
29. Barallobre, M. J. *et al.* Aberrant development of hippocampal circuits and altered neural activity in netrin 1-deficient mice. *Development* **127**, 4797–4810 (2000).
30. Deiner, M. S. *et al.* Netrin-1 and DCC mediate axon guidance locally at the optic disc: loss of function leads to optic nerve hypoplasia. *Neuron* **19**, 575–589 (1997).
31. Jiang, Y., Liu, M. T. & Gershon, M. D. Netrins and DCC in the guidance of migrating neural crest-derived cells in the developing bowel and pancreas. *Dev. Biol.* **258**, 364–384 (2003).
32. Barrett, C. & Guthrie, S. Expression patterns of the netrin receptor UNC5H1 among developing motor neurons in the embryonic rat hindbrain. *Mech. Dev.* **106**, 163–166 (2001).
33. Liu, Y. *et al.* Novel role for netrins in regulating epithelial behavior during lung branching morphogenesis. *Curr. Biol.* **14**, 897–905 (2004).
34. Leighton, P. A. *et al.* Defining brain wiring patterns and mechanisms through gene trapping in mice. *Nature* **410**, 174–179 (2001).
35. Lawson, N. D. & Weinstein, B. M. Arteries and veins: making a difference with zebrafish. *Nature Rev. Genet.* **3**, 674–682 (2002).
36. Lawson, N. D. & Weinstein, B. M. *In vivo* imaging of embryonic vascular development using transgenic zebrafish. *Dev. Biol.* **248**, 307–318 (2002).
37. Lauderdale, J. D., Davis, N. M. & Kuwada, J. Y. Axon tracts correlate with netrin-1a expression in the zebrafish embryo. *Mol. Cell. Neurosci.* **9**, 293–313 (1997).
38. Corset, V. *et al.* Netrin-1-mediated axon outgrowth and cAMP production requires interaction with adenosine A2b receptor. *Nature* **407**, 747–750 (2000).
39. Shirasaki, R., Mirzayan, C., Tessier-Lavigne, M. & Murakami, F. Guidance of circumferentially growing axons by netrin-dependent and -independent floor plate chemotropism in the vertebrate brain. *Neuron* **17**, 1079–1088 (1996).
40. Koch, M. *et al.* A novel member of the netrin family, beta-netrin, shares homology with the beta chain of laminin: identification, expression, and functional characterization. *J. Cell Biol.* **151**, 221–234 (2000).
41. Llambi, F., Causeret, F., Bloch-Gallego, E. & Mehlen, P. Netrin-1 acts as a survival factor via its receptors UNC5H and DCC. *EMBO J.* **20**, 2715–2722 (2001).
42. Tanikawa, C., Matsuda, K., Fukuda, S., Nakamura, Y. & Arakawa, H. p53RDL1 regulates p53-dependent apoptosis. *Nature Cell Biol.* **5**, 216–223 (2003).
43. Carmeliet, P. *et al.* Targeted deficiency or cytosolic truncation of the VE-cadherin gene in mice impairs VEGF-mediated endothelial survival and angiogenesis. *Cell* **98**, 147–157 (1999).
44. Yin, Y., Sanes, J. R. & Miner, J. H. Identification and expression of mouse netrin-4. *Mech. Dev.* **96**, 115–119 (2000).
45. Wang, H., Copeland, N. G., Gilbert, D. J., Jenkins, N. A. & Tessier-Lavigne, M. Netrin-3, a mouse homolog of human NTN2L, is highly expressed in sensory ganglia and shows differential binding to netrin receptors. *J. Neurosci.* **19**, 4938–4947 (1999).
46. Westerfield, M. *The Zebrafish Book. A Guide for the Laboratory Use of Zebrafish* (Danio Rerio) (Univ. Oregon Press, Eugene, 1994).
47. Stalmans, I. *et al.* VEGF: a modifier of the del22q11 (DiGeorge) syndrome? *Nature Med.* **9**, 173–182 (2003).
48. Sugiyama, D. *et al.* Erythropoiesis from acetyl LDL incorporating endothelial cells at the prelever stage. *Blood* **101**, 4733–4738 (2003).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank members of U36 for critical discussions; E. Etienne, A. Réaux, C. Esguerra and S. Maity for their expertise; C. Betsholtz for the *Pdgfrb* probe; H. Gerhardt for sharing unpublished data; and H. Rayburn, C. Jolicœur, J. Zhong, S. Terclavers, A. Claes, A. Vanhuffelen and S. Wyns for technical support. This work was supported by grants from Inserm (Avenir), Fondation Schlumberger pour l'Éducation et la Recherche (FSER), Ministère de l'Éducation et de la Recherche (ACI Biologie du Développement), Association pour la Recherche contre le Cancer (ARC) to A.E., a European Union grant to A.E. and P.C., and grants to M.T.-L. from the NIMH and the NHLBI. E.L.N. was supported by the Royal Dutch Academy of Science (KNAW) and Avenir; X.L. by the Damon Runyon Cancer Research Foundation; and M.T.-L. by the Howard Hughes Medical Institute.

Competing interests statement The authors declare competing financial interests: details accompany the paper on www.nature.com/nature.

Correspondence and requests for materials should be addressed to A.E. (anne.eichmann@college-de-france.fr) or M.T.-L. (tessier-lavigne.marc@gene.com). Zebrafish *Unc5b* is deposited in GenBank under accession number AY764274.

Crystal structure of RecBCD enzyme reveals a machine for processing DNA breaks

Martin R. Singleton^{1*}, Mark S. Dillingham^{2*}, Martin Gaudier¹, Stephen C. Kowalczykowski³ & Dale B. Wigley¹

¹Cancer Research UK Clare Hall Laboratories, The London Research Institute, Blanche Lane, South Mimms, Potters Bar, Herts. EN6 3LD, UK

²National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA, UK

³Sections of Microbiology and Molecular & Cellular Biology, Center for Genetics and Development, University of California at Davis, Davis, California 95616, USA

* These authors contributed equally to this work

RecBCD is a multi-functional enzyme complex that processes DNA ends resulting from a double-strand break. RecBCD is a bipolar helicase that splits the duplex into its component strands and digests them until encountering a recombinational hotspot (Chi site). The nuclease activity is then attenuated and RecBCD loads RecA onto the 3' tail of the DNA. Here we present the crystal structure of RecBCD bound to a DNA substrate. In this initiation complex, the DNA duplex has been split across the RecC subunit to create a fork with the separated strands each heading towards different helicase motor subunits. The strands pass along tunnels within the complex, both emerging adjacent to the nuclease domain of RecB. Passage of the 3' tail through one of these tunnels provides a mechanism for the recognition of a Chi sequence by RecC within the context of double-stranded DNA. Gating of this tunnel suggests how nuclease activity might be regulated.

Double-strand breaks in DNA are potentially lethal to a cell if not repaired. They arise in several ways, including DNA damage from ionizing radiation and as a result of single-strand nicks that produce a double-strand break as a replication fork passes by them (reviewed in ref. 1).

In eubacteria, double-strand breaks are repaired mainly by the homologous recombination pathway in a process initiated by the RecBCD/AddAB family of enzymes¹. RecBCD acts in a complicated way using a variety of different enzyme activities that are associated with the complex (Fig. 1). A combination of helicase and nuclease activities digests away the DNA until a Chi ('crossover hotspot instigator') site is encountered. Chi has the sequence 5'-GCTGGTGG-3' and is recognized as single-stranded DNA (ssDNA) but within a double-stranded context (that is, as the duplex is unwound²⁻⁴). At this point the protein pauses⁵ and the nuclease activity is both reduced⁶ and its polarity switched⁷. The final cleavage event on the 3' tail occurs at or within a few bases to the 3' side of Chi⁸, and the 3' tail is then protected from further digestion. Finally, RecBCD enzyme loads RecA protein onto the 3' tail to initiate recombination by means of the RecA pathway⁹.

The RecBCD enzyme comprises three subunits arranged as a 330-kDa heterotrimer that is fully functional without the need for further oligomerization¹⁰. The extraordinary range of enzyme activities catalysed by RecBCD can be attributed to the three subunits as follows: RecB is a 3'-5' helicase and multi-functional nuclease^{11,12}, RecC recognizes Chi¹³ and RecD is a 5'-3' helicase^{14,15}.

To gain a better understanding of the multifaceted mechanism of this DNA-processing machine we have determined the crystal structure of a complex of *Escherichia coli* RecBCD enzyme bound to a blunt-ended DNA hairpin. In addition to revealing the architecture of the enzyme, the structure shows how the two motor activities and the nuclease are coupled, allowing the recognition of Chi within the context of double-stranded DNA and suggesting a mechanism by which this regulates the polarity of the nuclease activity.

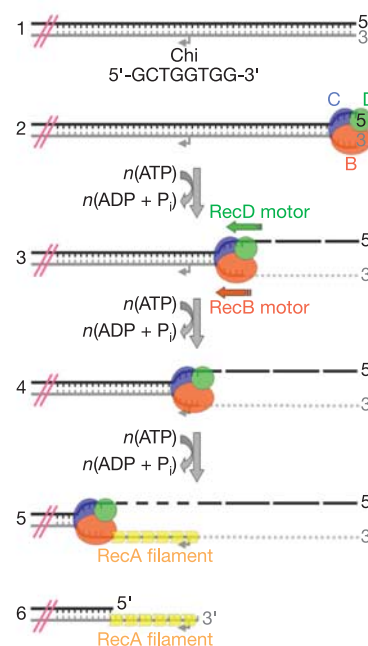


Figure 1 The processing of double-strand breaks by RecBCD enzyme. RecB is coloured orange, RecC is blue and RecD is green. In stage 1, a double-strand break is created as a result of DNA damage or a collapsed replication fork. The location of the eight base Chi site (sequence as shown) is represented by a bent arrow. In stage 2, RecBCD binds to the end of the DNA duplex and initiates unwinding. In stage 3, ATP-dependent DNA unwinding progresses and is coupled to the endonucleolytic digestion of both DNA strands. Although the 3' tail is cleaved frequently, the 5' tail is cleaved much less often. In stage 4, on encountering a Chi site in the 3' terminated strand, the enzyme pauses and digestion of the 3' tail ceases. In stage 5, cleavage of the 5' tail becomes more frequent and RecA protein is loaded on the 3' tail. In stage 6, RecBCD protein dissociates leaving a RecA-coated 3' tail that can initiate homologous recombination.

Structure of the bound DNA

RecBCD has a high affinity for blunt DNA ends¹⁶ with a footprint extending about 20 base pairs¹⁷. Consequently, we designed a self-complementary 43-mer oligonucleotide that forms a 19-base pair duplex with a five-base hairpin turn at one end and a blunt end at the other. As predicted, the DNA is bound with the blunt end of the duplex into the RecBCD complex. Remarkably, although the first 15 base pairs of the duplex from the hairpin form regular B-form DNA, the last four base pairs at the blunt end have been unwound by the protein. Experiments involving the chemical modification of RecBCD–DNA complexes have shown that DNA is unwound in the initiation complex even in the absence of ATP¹⁸.

RecB protein: endonuclease and 3′–5′ helicase

The RecB subunit is a helicase with 3′–5′ directionality¹⁹. Sequence analysis has identified RecB as a member of the helicase Superfamily 1 (SF1)²⁰ in common with several other helicases, such as PcrA and Rep, whose structures have been determined^{21,22}. Accordingly, the structure of the amino-terminal region of the RecB subunit is similar to that of other SF1 helicases (Fig. 2a) with regions equivalent to the canonical 1A, 1B, 2A and 2B domains. The structures of all four of these domains are very similar to their equivalents in PcrA (with a root-mean-square deviation (r.m.s.d.) for each domain of 1.9–3.2 Å for equivalent C α positions) but domain 1B contains an additional large insertion compared with PcrA. These residues form a discrete ‘arm’ structure that extends

from the surface of the complex alongside the duplex DNA substrate. The relative orientation of domains 1A and 2A is similar to that expected for the enzyme in the absence of ATP²³. However, domains 1B and 2B are in orientations that are quite different from those of any of the PcrA or Rep structures and seem to fulfil very different roles. For example, domain 2B of RecB forms a major interface with the RecC protein (see below) but this domain contacts duplex DNA in PcrA.

The 3′ tail of the bound DNA runs across domain 2A of the RecB subunit with the last three bases in a similar conformation to that seen for the equivalent residues in the ssDNA tail in the PcrA–DNA complex²³. Domains 1A and 2A are the canonical helicase motor domains responsible for the ATP-dependent translocation of the protein along ssDNA^{23–25}. Similar motor domains are found in both SF1 and SF2 helicases²⁶.

Domain 3, the RecB nuclease domain, is connected to the remainder of the protein by a long linker region of about 70 amino acids. This linker region has been shown to be sensitive to proteolytic cleavage¹². As suggested previously²⁷, the fold of this domain is similar to that of the core of the λ exonuclease²⁸ (r.m.s.d. of 3.5 Å for C α positions). Residue Asp 1080, known to be important for nuclease activity²⁹, is located at a position equivalent to the active site of λ exonuclease. There are three conserved acidic residues and a lysine residue in the active site (Glu 1020, Asp 1067, Asp 1080 and Lys 1082). Electron density that we attribute to a bound calcium ion from the crystallization medium is also

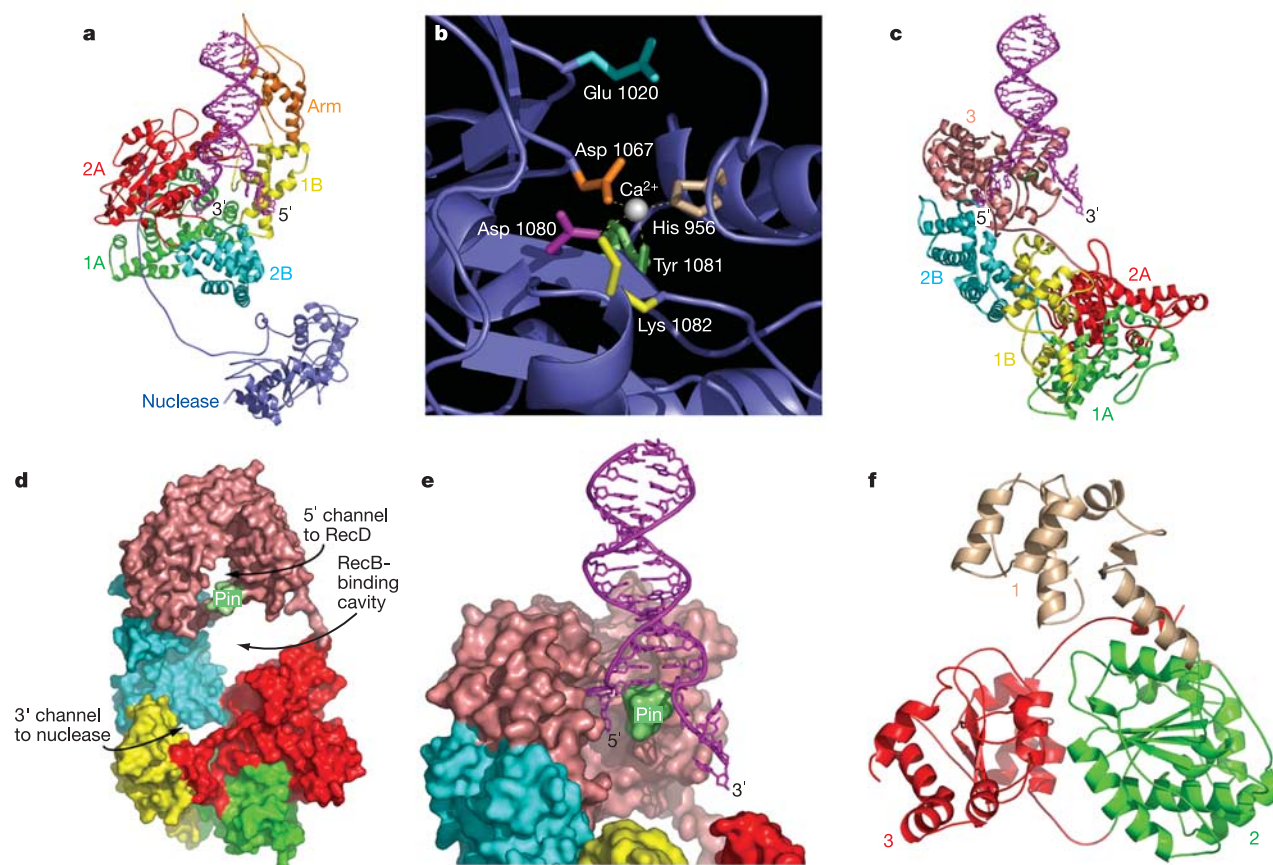


Figure 2 Structures of the individual RecBCD subunits. **a**, Domain structure of the RecB subunit. **b**, Close-up of the active site of the nuclease. The calcium ion (grey sphere) is coordinated to the side chains of three residues (His 956, Asp 1067 and Asp 1080) and the main-chain carbonyl of Tyr 1081. **c**, Domain structure of the RecC subunit. The pin region is highlighted. **d**, Space-filling representation of RecC, showing the channels

through the protein in the same colour scheme as in **c**. **e**, Close-up view of the pin region in RecC, showing how the DNA duplex is split across this feature of the RecC protein. **f**, Domain structure of the RecD subunit. Domains 2 and 3 are equivalent to the canonical 1A and 2A domains of other SF1 helicases. The images in **a**, and also those in Figs 2 and 4, were created with PyMOL (<http://www.pymol.org>).

present in this region (Fig. 2b). It is likely that this calcium ion is bound at the position where a magnesium ion would normally bind in the active site of the protein, because calcium has been shown to be an inhibitor of the nuclease activity³⁰.

RecC protein: duplex splitting and recognition of Chi

The overall fold of RecC is the most remarkable of the subunits (Fig. 2c) and contains three large channels that run through the protein (Fig. 2d). The largest of these accommodates the 2B domain of RecB and provides a major interface between the proteins. The other two channels are pathways along which the single-stranded tails of the DNA run to, or from, the two helicase subunits (see below). The protein can be divided into three domains. Quite unexpectedly, domains 1 and 2 have the same fold as canonical SF1 helicases (r.m.s.d. ~ 3.5 Å for equivalent C α positions of PcrA), despite a lack of sequence similarity including any of the characteristic helicase motifs. This strongly suggests that the RecC subunit evolved from a helicase ancestor but has now lost this helicase function. Furthermore, the similarity to helicase domains suggests a potential ssDNA-binding site that could provide a mechanism for the recognition of Chi as it passes through this site (see below).

Domain 3 of RecC is connected to the rest of the protein by a 20-residue linker that also forms the sides of the channel into which domain 2B of RecB binds (Fig. 3a). The fold of this domain has not previously been observed, according to the DALI server³¹. Domain 3 makes intimate contacts with each of the two separated strands of the DNA substrate running in different directions either side of a 'pin' that protrudes from the surface of the protein and serves to split the duplex (Fig. 2e). The 5' tail passes through another channel in RecC and heads towards the motor domains of RecD.

The open structure of RecC explains why it is readily proteolysed in solution. The major cleavage site of RecC is located after residue 804; this would release the C-terminal domain 3 of the protein, as suggested previously³². Deleting this domain eliminates RecD assembly within the RecBCD complex and this seems to relate to a small area of contacts with domain 2 of RecD, although the major interface between these proteins is between domain 1 of RecD and subdomain 2B of RecC (Fig. 3).

RecD protein: 5'–3' helicase

Although identified as a ssDNA-dependent ATPase several years

ago³³ and shown to contain several characteristic helicase motifs²⁰, it was only recently shown that the RecD subunit has 5'–3' helicase activity^{14,15}. The structure of the RecD subunit is the first of a SF1 helicase with 5'–3' directionality. The protein comprises three domains (Fig. 2f). Domain 1 forms the principal interface between RecD and domain 2B of RecC. Domains 2 and 3 are similar to the motor domains (1A and 2A) of other SF1 helicases (r.m.s.d. 3.9 and 3.3 Å on C α positions, respectively). There is a disordered region of about 55 amino acids in domain 3, which would correspond to domain 2B of PcrA. The rest of domain 3 is also highly mobile, with higher *B* factors than for the rest of the structure.

Although most SF1 helicases have 3'–5' polarity, some (such as RecD and Dda) have the opposite polarity^{14,34}. There are two simple ways in which a SF1 helicase could be adapted to alter the directionality to 5'–3', either to reverse the direction of translocation along the bound DNA or to bind the DNA in the opposite orientation. The path taken by the bound ssDNA tail across RecC would present the 5' tail to the helicase domains in an orientation similar to that seen for other helicases, suggesting that the alteration in directionality of helicases in SF1 is based on the direction of movement along the bound DNA rather than the directionality of ssDNA binding itself. It has been shown that residues in motif 1a have a pivotal function in directional translocation in SF1 helicases³⁵, and it is probably significant that these residues are absent from RecD.

The RecBCD complex

The structures of the individual subunits of RecBCD reveal why they are difficult to express and prepare, unless complexed together. The RecB and RecC proteins, in particular, are wrapped tightly around one another. In isolation, both proteins have exposed linker regions that are susceptible to proteolysis but protected when the proteins are bound to each other. This intimate embrace of the proteins is essential for their function and allows the ssDNA strands to be shepherded through the enzyme, providing a temporal link between their separation and subsequent endonucleolytic digestion.

The bound DNA makes extensive contacts with the RecB and RecC subunits. The 'arm' structure of the RecB protein interacts with duplex DNA ahead of the fork about 12 base pairs from the junction. The RecC protein contacts both strands of DNA and splits them before feeding the 3' strand to the RecB protein and the 5' tail

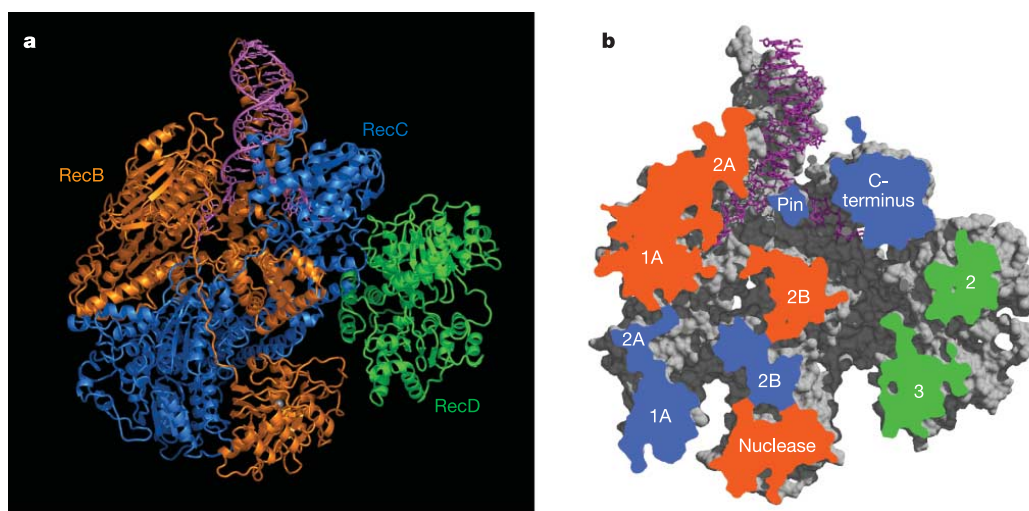


Figure 3 Structure of the RecBCD–DNA complex. **a**, The entire RecBCD–DNA complex. The bound DNA is coloured magenta and the bound calcium ion is a grey sphere. **b**, Cutaway view showing the channels through the RecBCD complex. The cutaway surface of the RecB subunit is orange, that of RecC is blue and that of RecD is green. The

bound DNA is coloured magenta. Numbers refer to the domains of the appropriate subunits. Domains 2 and 3 of RecD are equivalent to canonical helicase domains 1A and 2A. Both **b** and Fig. 4a were created with MSMS⁴⁹ and rendered with Raster3D⁵⁰.

to the RecD subunit. Consequently, the protein complex covers the first 16 bases of the 5'-terminated strand of the DNA and 13 bases on the 3'-terminated strand. These results are consistent with DNase I footprinting studies of the initiation complex that showed protection of 20 or 21 bases on the 5' tail and 16 or 17 bases on the 3' tail¹⁷. Furthermore, these studies showed that ultraviolet radiation crosslinked the 3' tail to the RecB protein, which is consistent with the crystal structure. The 5' tail was crosslinked to both the RecC and RecD subunits. Although we observe extensive contacts between the RecC subunit and the bound DNA, there are no contacts with the RecD subunit in our structure, although, as unwinding of the DNA progresses, the 5' tail would be fed from the RecC subunit towards the motor domains of RecD. This large footprint also relates to experiments that examined the activity of RecBC on gapped duplex substrates³⁶. These experiments showed that RecBC can pass over single-stranded gaps of up to 23 bases provided that the 3' strand is intact. This property of the protein most probably relates to the arm structure of RecB that contacts duplex DNA ahead of the junction, allowing the protein to step over intervening single-strand regions.

RecBCD has both 3'–5' and 5'–3' helicase activities, contributed by the RecB and RecD subunits, respectively. These subunits drive DNA unwinding by acting as ssDNA motors, pulling the two antiparallel strands of the DNA across the pin of the RecC subunit and thus splitting the duplex. RecB is closely related to the well-studied SF1 helicase, PcrA. The N-terminal region of RecB is the same as PcrA with the exception of an insertion of about 115 residues within domain 1B (the 'arm') that contacts the duplex ahead of the junction in a manner similar to contacts in the PcrA–DNA complex²³, except that this contact is provided by the 2B domain in PcrA. In PcrA, the strength of this contact is dependent on ATP binding²⁵, and it may be that this is also true for RecBCD. The present structure was determined in the absence of ATP and the *B* factors of the residues in the arm are considerably higher than the average for the rest of the protein, suggesting that this part of the structure is more mobile. Furthermore, in PcrA, the duplex is split across a 'pin' that protrudes from domain 2B in a manner similar to the pin in RecBCD, except that in the latter case the pin is contributed by the RecC subunit. Consequently, we observe an apparent conservation of mechanism but contributed by different parts of the structure.

The two helicase motors can apparently work independently of

one another. Electron microscopy data have shown that loops can be extruded from the complex when the helicase activity of either subunit is inactivated¹⁵. A ssDNA loop is also extruded from the protein on encountering a Chi site⁵.

Isolated RecB protein has been shown to have weak 3'–5' helicase activity¹⁹, but the processivity of this activity is greatly enhanced when complexed with RecC¹². The crystal structure reveals how RecC contributes to the helicase activity both directly, through the pin that splits the DNA duplex at the junction, and indirectly, by sequestering the 3' tail of the substrate and preventing reannealing. Passage of the DNA through the complex would also provide a steric block to dissociation of the protein, a tactic frequently adopted by other processive enzymes in DNA metabolism.

PcrA has been shown to use one ATP molecule for each base pair that it translocates and, by inference, during helicase action as well²⁴. By contrast, RecBCD uses between two and three ATP per base pair separated^{16,37} although RecBC, or RecBCD containing an ATPase-deficient RecD subunit, uses only one ATP per base pair³⁷. This suggests that one ATP molecule is used by each helicase motor during translocation, a decrease in efficiency that is apparently tolerated in exchange for the increased processivity of RecBCD in comparison with RecBC, or the ability of the holoenzyme to bypass single-stranded gaps.

The discovery that RecC might be a defunct helicase raises some important implications about the interaction of DNA with the complex. The channel for the 3'-terminated strand passes from RecC, across the RecB motor domains, then back through RecC until it emerges adjacent to the nuclease domain of RecB (Fig. 3b). In passing through RecC, the channel runs across the top of the 'helicase' domains, which corresponds to the ssDNA-binding site in helicases. Consistent with this proposal is the observation that the channel is lined with conserved basic and exposed hydrophobic residues, as would be expected for a site that binds ssDNA. The ssDNA-binding site in SF1 helicases spans about eight bases^{22,23}, and it is probably significant that the Chi sequence is also eight bases in length (5'-GCTGGTGG-3'). Recognition of ssDNA by helicases is non-specific with regard to sequence. This is achieved by providing complementary charge for the phosphodiester backbone and exposed aromatic residues to stack with the DNA bases. Contacts with the functional groups on the bases are avoided, because this would necessarily result in unwanted sequence specificity. However, the introduction of such contacts could transform a canonical

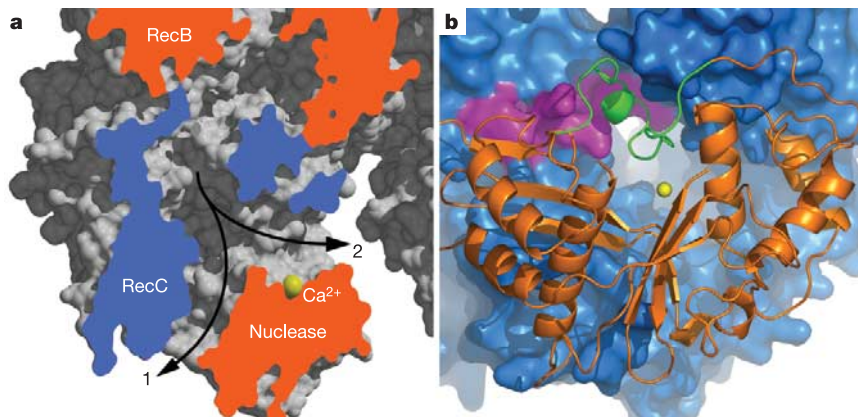


Figure 4 Alternative exits from the 3' tunnel. **a**, Cutaway view of the exit channels running each side of the nuclease domain. The calcium ion at the nuclease active site is coloured yellow. For the purposes of this figure, the loop that blocks the channel has been omitted. There are two exit channels from the RecC subunit. One of these (labelled 1) bypasses the nuclease site. Access to the nuclease active site through channel 2 is blocked by a helix in the structure. **b**, The interface between the RecC subunit and the

RecB nuclease domain viewed across the RecB nuclease active site. RecC is shown in blue as a space-filling representation, with the region affected in the RecC* mutants highlighted in magenta. The RecB nuclease domain is overlaid as an orange ribbon and the bound calcium ion as a yellow sphere. Access to the nuclease active site from the channel is blocked by a loop from the nuclease domain that includes an α -helix (residues 909–930, coloured green).

helicase ssDNA translocation site into a ssDNA ‘scanning’ site that would provide a convenient mechanism by which RecC could recognize a Chi sequence as it passed through the channel.

There is a group of RecC mutants with altered Chi recognition^{38–40}. The mutations (residues 647–663) map to a region at the opening of the channel that accommodates the 3′ tail, at the interface with the nuclease domain of RecB (Fig. 4a). It has been shown that on Chi recognition the last DNA cleavage event is no more than four to six bases to the 3′ side of the Chi site⁸. This suggests that the end of the Chi recognition site should be 15–25 Å from the nuclease active site depending on the degree of extension of the ssDNA strand. Residues 647–663 are located 20–25 Å from the acidic triad in the nuclease domain. These mutations are frameshifts across several residues rather than point mutations, mostly introducing deletions in the sequence with undetermined consequences on the local structure. Furthermore, this region is the linker between the 2A and 2B domains, so mutations here could have significant consequences on the overall conformation of the protein. It is therefore difficult to draw detailed information from these mutants other than identifying this region as important for Chi recognition. The

complex pathway taken by the ssDNA suggests a mechanism for the observation that Chi sequences are recognized as ssDNA but arising from double-stranded DNA that is unwound by the enzyme^{2–4}. During unwinding, the channels, and hence also the Chi-binding site, would be occupied with ssDNA, thus blocking the binding of oligonucleotides containing Chi sequences when added *in trans* to the substrate after DNA unwinding has begun⁴¹.

Regulation of nuclease activities

The nuclease activities of RecB are carefully regulated in several ways, including switching cleavage between 5′ and 3′ tails, 3′ nuclease attenuation after Chi, and nuclease activation by RecD. The location of the nuclease active site allows processive hydrolysis of the 3′ tail as it emerges from the RecC subunit. However, proximity of the 5′ tail would enable it to compete with the 3′ tail for binding at the nuclease, although this would be less frequent owing to the less favourable disposition of the 5′ tail, which would access the site from the opposite direction. Nuclease digestion of the 3′ tail would be attenuated after Chi simply as a consequence of binding tightly to the RecC subunit, resulting in more frequent cleavage of the 5′ tail resulting from less hindered access. Although we do not need to invoke a conformational change in the protein after it has encountered Chi, we cannot exclude this possibility.

Interestingly, there are two exits from the 3′ tunnel as it emerges from RecC (Fig. 4a). One of these passes along the back of the nuclease domain, bypassing the nuclease active site. The other is blocked by an α-helix from RecB (Fig. 4b). This would provide a simple mechanism for controlling the nuclease activity of the complex such that access to the active site would be blocked if this helix remained as positioned in this structure, forcing the strand to exit without passing through the nuclease. The helix is connected to the rest of the protein by two long flexible linkers, suggesting that it might be able to swing out of the way. The RecD subunit is involved in stimulating the nuclease activity⁴² and, although the trigger for movement of this helix remains unclear, domain 3 of RecD is in close proximity.

A mechanism for processing DNA ends

The crystal structure allows us to understand RecBCD activity in better molecular detail (Fig. 5). Once bound to a blunt end, the initiation complex is formed in which the double strands of the duplex are opened up and fed towards the motor subunits. The DNA is unwound by the combined activities of the two helicase subunits

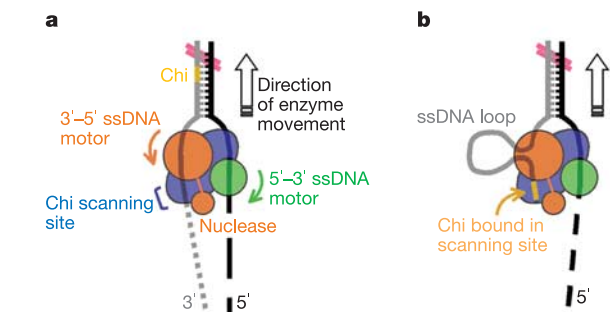


Figure 5 Diagram outlining the changes in RecBCD that occur after encountering a Chi site. **a**, Before Chi, the enzyme progresses along duplex DNA using the bipolar motors of the RecB (orange) and RecD (green) subunits. The 3′ tail is digested processively but the 5′ tail is cut much less frequently. **b**, On encountering Chi, the RecC subunit (blue) binds tightly to the 3′ tail, preventing further digestion of this strand. The 5′ tail is now able to access the nuclease site more frequently and is degraded more fully. The enzyme continues to advance along the DNA resulting in a loop out from the RecB subunit that can be loaded with RecA protein.

Table 1 Crystallographic statistics

Data collection	Native	TaBr1	TaBr2	Se1	IodoU
Resolution (Å)	30–3.1	30–3.8	30–5.5	20–3.2	30–3.1
Completeness (%)	99.8	91.4	91.3	99.5	95.8
R_{symm} (%)	7.1	14.9	9.3	11.3	8.5
R_{deriv} (%)	–	16.2	22.3	13.5	17.4
No. sites	–	10	10	119	18
Phasing power	–	0.32	0.97	0.60	0.28
Anomalous data statistics	Se1 (peak)	Se2 (inflection)	Se3 (remote)	TaBr3 (peak)	
Resolution (Å)	30–3.2	30–3.4	30–3.6	20–5.0	
Wavelength (Å)	0.97942	0.97960	0.93927	1.25485	
Completeness (%)	99.5	99.7	99.7	90.3	
R_{symm} (%)	11.3	11.9	14.3	6.8	
Phasing power	0.44	0.63	0.27	–	
Final model	Value				
R factor (All data excluding free R set) (%)	24.2				
R_{free} (5% of data) (%)	29.6				
R.m.s.d. bond length (Å)	0.018				
R.m.s.d. bond angle (deg)	2.16				

that pull the DNA strands across the pin of RecC, thus splitting the duplex. The 5' tail is fed to the RecD helicase and then onto the nuclease domain of RecB for digestion. The 3' tail is fed along a channel within the protein complex that emerges at the nuclease active site. Consequently, this strand is digested processively as the protein complex progresses along the DNA. The 5' tail is also digested, but less frequently because it is not situated as favourably as the 3' tail and cannot compete as effectively for the nuclease active site. As the 3' tail is fed through the RecC subunit it is able to scan the DNA for a Chi sequence. On encountering Chi, the RecC subunit binds tightly to this site in the 3' tail, thus preventing further digestion of this strand, with the final cleavage event taking place at, or close to, the Chi site. The 5' tail is now able to access the nuclease site more readily and is therefore cleaved more frequently. Finally, RecBCD loads RecA protein onto the 3' tail by a mechanism that is not yet fully understood but involves the nuclease domain of RecB⁴³ before dissociating from the DNA. □

Methods

Protein expression and purification, and DNA preparation

Full-length RecB, RecC and RecD were expressed together with the use of the pPB520 and pPB800 plasmids¹¹ in a *ΔrecBCD E. coli* strain (V330) containing the LacI^q overexpression plasmid pMS421. The RecBCD complex was purified by ammonium sulphate fractionation (0–50% w/v), followed by Fast Flow Q (Amersham) anion-exchange chromatography. RecBCD-containing fractions were dialysed overnight and further purified by heparin–Sepharose (Amersham), Mono-Q (Amersham) and gel-filtration (Superdex 200; Amersham) chromatography. The DNA hairpin was prepared by heating the self-complementary oligonucleotide (5'-TCTAATGCGAGCACTGCTATTCCTAG CAGTGTCTCGCATAG-3') to 95 °C for 5 min followed by rapid cooling in ice. The DNA was then purified by anion-exchange chromatography and desalted by gel filtration.

Crystallization and structure determination

Protein was concentrated to 10 mg ml⁻¹ in 10 mM Tris-HCl pH 7.5, 100 mM NaCl and 1 mM dithiothreitol, with DNA present in a 1.3:1 molar excess. Hanging drops were set up above reservoir solutions of 100 mM HEPES pH 7.0, 200 mM calcium acetate, 4–8% (w/v) poly(ethylene glycol) 2000. Crystal nucleation was initiated by microseeding, and crystals grew to a full size of about 400 × 100 × 60 μm³ within 5 days.

Crystals were cryoprotected by rapid sequential transfer through the reservoir solution with the addition of ethylene glycol in 5% (v/v) steps to a final concentration of 30%. Diffraction data were collected on beamlines 14.1 and 14.4 at the European Synchrotron Radiation Facility, Grenoble. The crystals were of space group P2₁2₁2₁ with unit cell dimensions *a* = 133 Å, *b* = 187 Å, *c* = 335 Å and contained two RecBCD–DNA complexes in the asymmetric unit.

Diffraction data (Table 1) were integrated and scaled together with the XDS package⁴⁴. The positions of ten Ta₆Br₁₂ clusters were determined by analysis of anomalous diffraction data collected at the Ta L_{III} edge (TaBr3) using the program Shake 'n Bake⁴⁵. However, these data were too non-isomorphous to be useful for multiple isomorphous replacement (MIR) phasing with native data. Consequently, data from two lower-occupancy soaks were collected and used for MIR phasing (TaBr1 and TaBr2). Phases calculated from these data sets allowed the location of 18 iodine sites in data from a crystal grown with iodouracil-substituted oligonucleotide and 119 selenium sites in selenomethionine-substituted protein. Heavy-atom parameters were refined and phases calculated with the program SHARP⁴⁶, using a combination of two runs, one with regular MIRAS phasing (including the data from Se1 as a derivative and using a spherically averaged approximation for the tantalum clusters) and the other with multiwavelength anomalous diffraction phasing for the three selenium data sets collected from the same crystal (Se1, Se2 and Se3). Phases from each run were combined, resulting in a set of phase estimates with an overall mean figure of merit of 0.33. These initial phases were improved by density modification with twofold non-crystallographic symmetry (NCS) averaging in DM⁴⁷. These maps permitted location of the DNA and building of most of one complex. Initial refinement was performed by using the CNS package⁴⁸ with NCS constraints. Further rounds of model building were undertaken between refinement cycles. Towards the end of refinement the NCS constraints were relaxed and parts of the structure that did not obey local symmetry were refined independently. The final model has an *R* factor of 24.2% (*R*_{free} = 29.6%).

Received 1 July; accepted 2 September 2004; doi:10.1038/nature02988.

- Kowalczykowski, S. C. Initiation of genetic recombination and recombination-dependent replication. *Trends Biochem. Sci.* **25**, 156–165 (2000).
- Ponticelli, A. S., Schultz, D. W., Taylor, A. F. & Smith, G. R. Chi-dependent DNA strand cleavage by the RecBC enzyme. *Cell* **41**, 145–151 (1985).
- Taylor, A. F., Schultz, D. W., Ponticelli, A. S. & Smith, G. R. RecBC enzyme nicking at Chi sites during DNA unwinding: location and orientation-dependence of the cutting. *Cell* **41**, 153–163 (1985).
- Bianco, P. R. & Kowalczykowski, S. C. The recombination hotspot Chi is recognized by the translocating RecBCD enzyme as the single strand of DNA containing the sequence 5'-GCTGTGG-3'. *Proc. Natl Acad. Sci. USA* **94**, 6706–6711 (1997).
- Spies, M. *et al.* A molecular throttle: the recombination hotspot Chi controls DNA translocation by the RecBCD helicase. *Cell* **114**, 647–654 (2003).

- Dixon, D. A. & Kowalczykowski, S. C. The recombination hotspot Chi is a regulatory sequence that acts by attenuating the nuclease activity of the *E. coli* RecBCD enzyme. *Cell* **73**, 87–96 (1993).
- Anderson, D. G. & Kowalczykowski, S. C. The recombination hot spot Chi is a regulatory element that switches the polarity of DNA degradation by the RecBCD enzyme. *Genes Dev.* **11**, 571–581 (1997a).
- Taylor, A. F. & Smith, G. R. Strand specificity of nicking of DNA at Chi sites by RecBCD enzyme. *J. Biol. Chem.* **270**, 24459–24467 (1995b).
- Anderson, D. G. & Kowalczykowski, S. C. The translocating RecBCD enzyme stimulates recombination by directing RecA protein onto ssDNA in a Chi-regulated manner. *Cell* **90**, 77–86 (1997b).
- Taylor, A. F. & Smith, G. R. Monomeric RecBCD enzyme binds and unwinds DNA. *J. Biol. Chem.* **270**, 24451–24458 (1995a).
- Boehmer, P. E. & Emmerson, P. T. *Escherichia coli* RecBCD enzyme: inducible overproduction and reconstitution of the ATP-dependent deoxyribonuclease from purified subunits. *Gene* **102**, 1–6 (1991).
- Yu, M., Souaya, J. & Julin, D. A. The 30 kDa C-terminal domain of the RecB protein is critical for the nuclease activity, but not the helicase activity, of the RecBCD enzyme from *Escherichia coli*. *Proc. Natl Acad. Sci. USA* **95**, 981–986 (1998).
- Handa, N., Ohashi, S., Kusano, K. & Kobayashi, I. Chi*, a chi-related 11-mer sequence partially active in an *E. coli* recC1004 strain. *Genes Cells* **2**, 525–536 (1997).
- Dillingham, M. S., Spies, M. & Kowalczykowski, S. C. RecBCD enzyme is a bipolar DNA helicase. *Nature* **423**, 893–897 (2003).
- Taylor, A. F. & Smith, G. R. RecBCD enzyme is a DNA helicase with fast and slow motors of opposite polarity. *Nature* **423**, 889–893 (2003).
- Roman, L. J. & Kowalczykowski, S. C. Characterisation of the adenosinetriphosphatase activity of the *Escherichia coli* RecBCD enzyme: relationship of ATP hydrolysis to the unwinding of duplex DNA. *Biochemistry* **28**, 2873–2881 (1989).
- Ganesan, S. & Smith, G. R. Strand-specific binding to duplex DNA ends by the subunits of the *Escherichia coli* RecBCD enzyme. *J. Mol. Biol.* **229**, 67–78 (1993).
- Farah, J. A. & Smith, G. R. The RecBCD enzyme initiation complex for DNA unwinding: enzyme positioning and DNA opening. *J. Mol. Biol.* **272**, 699–715 (1997).
- Boehmer, P. E. & Emmerson, P. T. The RecB subunit of the *Escherichia coli* RecBCD enzyme couples ATP hydrolysis to DNA unwinding. *J. Biol. Chem.* **267**, 4981–4987 (1992).
- Gorbalenya, A. E. & Koonin, E. V. Helicases: amino acid sequence comparisons and structure-function relationships. *Curr. Opin. Struct. Biol.* **3**, 419–429 (1993).
- Subramanya, H. S., Bird, L. E., Brannigan, J. A. & Wigley, D. B. Crystal structure of a DExx box helicase. *Nature* **384**, 379–383 (1996).
- Korolev, S., Hsieh, J., Gauss, G. H., Lohman, T. M. & Waksman, G. Major domain swiveling revealed by the crystal structures of complexes of *E. coli* Rep helicase bound to single-stranded DNA and ADP. *Cell* **90**, 635–647 (1997).
- Velankar, S. S., Soultanas, P., Dillingham, M. S., Subramanya, H. S. & Wigley, D. B. Crystal structures of complexes of PcrA helicase with a DNA substrate indicate an inchworm mechanism. *Cell* **97**, 75–84 (1999).
- Dillingham, M. S., Wigley, D. B. & Webb, M. R. Unidirectional single-stranded DNA translocation by PcrA helicase: measurement of step size and translocation speed. *Biochemistry* **39**, 205–212 (2000).
- Soultanas, P., Dillingham, M. S., Wiley, P., Webb, M. R. & Wigley, D. B. Uncoupling DNA translocation and helicase activity in PcrA: direct evidence for an active mechanism. *EMBO J.* **19**, 3799–3810 (2000).
- Singleton, M. R. & Wigley, D. B. Modularity and specialisation in Superfamily 1 and 2 helicases. *J. Bacteriol.* **184**, 1819–1826 (2002).
- Aravind, L., Makarova, K. S. & Koonin, E. V. Holliday junction resolvases and related nucleases: identification of new families, phyletic distribution and evolutionary trajectories. *Nucleic Acids Res.* **28**, 3417–3432 (2000).
- Kovall, R. & Matthews, B. W. Toroidal structure of λ-exonuclease. *Science* **277**, 1824–1827 (1997).
- Yu, M., Souaya, J. & Julin, D. A. Identification of the nuclease active site in the multifunctional RecBCD enzyme by creation of a chimeric enzyme. *J. Mol. Biol.* **283**, 797–808 (1998).
- Rosamond, J., Telander, K. M. & Linn, S. Modulation of the action of the RecBC enzyme of *Escherichia coli* K-12 by Ca²⁺. *J. Biol. Chem.* **254**, 8648–8652 (1979).
- Holm, L. & Sander, C. Protein structure comparison by alignment of distance matrices. *J. Mol. Biol.* **233**, 123–138 (1993).
- Amundsen, S. K., Taylor, A. F. & Smith, G. R. Domain of RecC required for assembly of the regulatory RecD subunit into the *Escherichia coli* RecBCD holoenzyme. *Genetics* **161**, 483–492 (2002).
- Chen, H. W., Ruan, B., Yu, M., Wang, J. & Julin, D. A. The RecD subunit of the RecBCD enzyme from *Escherichia coli* is a single-stranded DNA dependent ATPase. *J. Biol. Chem.* **272**, 10072–10079 (1997).
- Kuhn, B., Abdel-Monem, M., Krell, H. & Hoffmann-Berling, H. Evidence for two mechanisms for DNA unwinding catalyzed by DNA helicases. *J. Biol. Chem.* **254**, 11343–11350 (1979).
- Dillingham, M. S., Soultanas, P., Wiley, P., Webb, M. R. & Wigley, D. B. Defining the roles of individual residues in the single-stranded DNA binding site of PcrA helicase. *Proc. Natl Acad. Sci. USA* **98**, 8381–8387 (2001).
- Bianco, P. R. & Kowalczykowski, S. C. Translocation step size and mechanism of the RecBC DNA helicase. *Nature* **405**, 368–372 (2000).
- Korang, F. & Julin, D. A. Efficiency of ATP hydrolysis and DNA unwinding by the RecBC enzyme from *Escherichia coli*. *Biochemistry* **33**, 9552–9560 (1994).
- Schultz, D. W., Taylor, A. F. & Smith, G. R. *Escherichia coli* RecBC pseudorevertants lacking Chi recombinational hotspot activity. *J. Bacteriol.* **155**, 664–680 (1983).
- Arnold, D. A., Bianco, P. R. & Kowalczykowski, S. C. The reduced levels of Chi recognition exhibited by the RecBC^{1004D} enzyme reflect its recombination defect *in vivo*. *J. Biol. Chem.* **273**, 16476–16486 (1998).
- Arnold, D. A., Handa, N., Kobayashi, I. & Kowalczykowski, S. C. A novel, 11 nucleotide variant of Chi, Chi*: One of a class of sequences defining the *Escherichia coli* recombination hotspot Chi. *J. Mol. Biol.* **300**, 469–479 (2000).
- Kulkarni, A. & Julin, D. A. Specific inhibition of the *E. coli* RecBCD enzyme by Chi sequences in single-stranded oligodeoxynucleotides. *Nucleic Acids Res.* **32**, 3672–3682 (2004).
- Palas, K. M. & Kushner, S. R. Biochemical and physical properties of exonuclease V from *Escherichia coli*. *J. Biol. Chem.* **265**, 3447–3454 (1990).
- Churchill, J. J. & Kowalczykowski, S. C. Identification of the RecA protein-loading domain of RecBCD enzyme. *J. Mol. Biol.* **297**, 537–542 (2000).

44. Kabsch, W. Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J. Appl. Crystallogr.* **26**, 795–800 (1993).
45. Weeks, C. M. & Miller, R. Optimising Shake-and-bake for proteins. *Acta Crystallogr. D* **55**, 492–500 (1999).
46. de La Fortelle, E. & Bricogne, G. Maximum-likelihood heavy-atom parameter refinement for the multiple isomorphous replacement and multiwavelength anomalous diffraction methods. *Methods Enzymol.* **276**, 472–494 (1997).
47. Collaborative Computing Project No. 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
48. Brunger, A. T. *et al.* Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 904–925 (1998).
49. Sanner, M. F., Spehner, J. C. & Olson, A. J. Reduced surface: an efficient way to compute molecular surfaces. *Biopolymers* **38**, 305–320 (1996).
50. Merritt, E. A. & Bacon, D. J. Raster3D: Photorealistic molecular graphics. *Methods Enzymol.* **277**, 505–524 (1997).

Acknowledgements We thank V. Ramakrishnan and J. Lowe for advice on the use and supply of tantalum bromide clusters, S. Halford for discussions about nucleases, C. Flensburg for advice on using SHARP and a pre-release version of the program, V. Dearing for oligonucleotide synthesis and purification, and D. Hall and E. Gordon for time and assistance on ESRF beamlines. This work was supported by Cancer Research UK (D.B.W.), NIH (S.C.K.) and a Wellcome Trust Travelling Research Fellowship (M.S.D.).

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to D.W. (dale.wigley@cancer.org.uk). The coordinates have been deposited at the Protein Databank with accession code 1W36.

A characteristic size of ~ 10 Mpc for the ionized bubbles at the end of cosmic reionization

J. Stuart B. Wyithe¹ & Abraham Loeb²

¹School of Physics, The University of Melbourne, Parkville, Victoria 3010, Australia

²Astronomy Department, Harvard University, 60 Garden Street, Cambridge, Massachusetts 02138, USA

The first galaxies to appear in the Universe at redshifts $z > 20$ created ionized bubbles in the intergalactic medium of neutral hydrogen left over from the Big Bang. The ionized bubbles grew with time, surrounding clusters of dwarf galaxies^{1,2} and eventually overlapped quickly throughout the Universe over a narrow redshift interval near $z \approx 6$. This event signalled the end of the reionization epoch when the Universe was a billion years old. Measuring the size distribution of the bubbles at their final overlap phase is a focus of forthcoming programmes to observe highly redshifted radio emission from atomic hydrogen. Here we show that the combined constraints of cosmic variance and light travel time imply an observed bubble size at the end of the overlap epoch of ~ 10 physical Mpc, and a scatter in the observed redshift of overlap along different lines-of-sight of ~ 0.15 . This scatter is consistent with observational constraints from recent spectroscopic data on the farthest known quasars. This implies that future radio experiments should be tuned to a characteristic angular scale of 0.5 degrees and have a minimum frequency bandwidth of ~ 8 MHz for an optimal detection of 21-cm flux fluctuations near the end of reionization.

During the reionization epoch, the characteristic bubble size (defined here as the spherically averaged mean radius of the H II regions that contain most of the ionized volume²) increased with time as smaller bubbles combined until their overlap completed and the diffuse intergalactic medium (IGM) was reionized. However, the largest size of isolated bubbles (fully surrounded by H I boundaries) that can be observed is finite, because of the combined phenomena of cosmic variance and light travel time. Figure 1 presents a schematic illustration of the geometry. There is a surface on the sky corresponding to the time along different lines-of-sight when the diffuse (uncollapsed) IGM was most recently neutral. We refer to it as the surface of bubble overlap (SBO). There are two competing sources for fluctuations in the SBO, each of which is dependent on the characteristic size, R_{SBO} , of the ionized regions just before the final overlap. First, the finite speed of light implies that photons observed from different points along the curved boundary of an H II region must have been emitted at different times during the history of the Universe. Second, bubbles on a co-moving scale R achieve reionization over a spread of redshifts owing to cosmic variance in the initial conditions of the density field smoothed on that scale. The characteristic scale of H II bubbles grows with time, leading to a decline in the spread of their formation redshifts¹ as the cosmic variance is averaged over an increasing spatial volume. However, the light travel time across a bubble rises concurrently. Suppose a signal photon that encodes the presence of neutral gas (for example, a 21-cm-line photon) is emitted from the far edge of the ionizing bubble. If the adjacent region along the line-of-sight has not become ionized by the time this photon reaches the near side of the bubble, then the photon will encounter diffuse neutral gas. Other photons emitted at this lower redshift will therefore also encode the presence of diffuse neutral gas, implying that the first photon was emitted prior to overlap, and not from the

SBO. Hence the largest observable scale of H II regions when their overlap completes corresponds to the first epoch at which the light crossing time becomes larger than the spread in formation times of ionized regions. Only then will the signal photon leaving the far side of the H II region have the lowest redshift of any signal photon along that line-of-sight.

The observed spectra of all quasars beyond $z \approx 6.1$ each show a Gunn–Peterson trough^{3,4}, a blank spectral region at wavelengths shorter than Lyman- α ($\text{Ly}\alpha$) at the quasar redshift, indicating the presence of H I in the diffuse IGM. The detection of Gunn–Peterson troughs indicates a rapid change^{5–7} in the neutral content of the IGM at $z \approx 6$, and hence a rapid change in the intensity of the background ionizing flux. This rapid change implies that

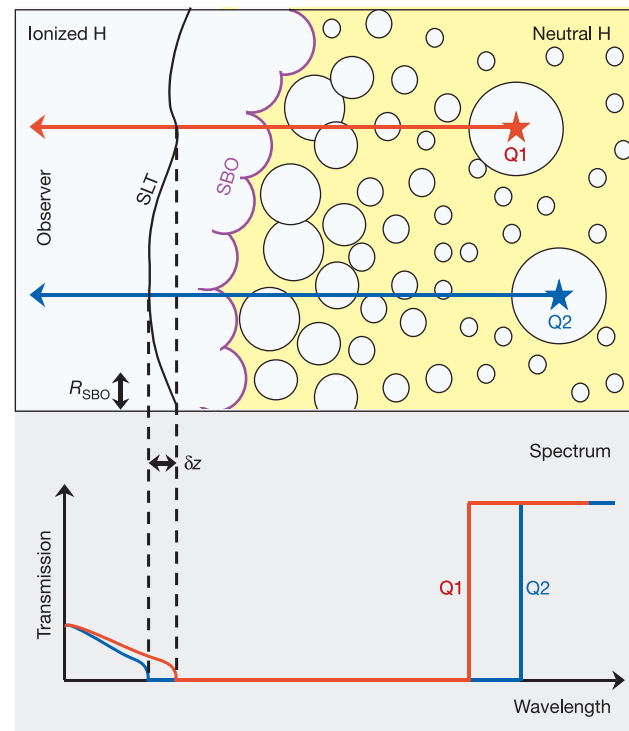


Figure 1 The distances to the observed surface of bubble overlap (SBO) and surface of Ly α transmission (SLT) fluctuate on the sky. The SBO corresponds to the first region of diffuse neutral IGM observed along a random line-of-sight. It fluctuates across a shell with a minimum width dictated by the condition that the light crossing time across the characteristic radius R_{SBO} of ionized bubbles equals the cosmic scatter in their formation times. Thus, light travel time and cosmic variance determine the characteristic scale of bubbles at the completion of bubble overlap. After some time delay, the IGM becomes transparent to Ly α photons, resulting in a second surface, the SLT. The upper panel illustrates how the lines-of-sight towards two quasars (Q1 in red and Q2 in blue) intersect the SLT with a redshift difference δz . The resulting variation in the observed spectrum of the two quasars is shown in the lower panel. Observationally, the ensemble of redshifts down to which the Gunn–Peterson troughs are seen in the spectra of $z > 6.1$ quasars is drawn from the probability distribution dP/dz_{SLT} for the redshift at which the IGM started to allow Ly α transmission along random lines-of-sight. The observed values of z_{SLT} show a small scatter⁴ in the SLT redshift around an average value of $\langle z_{\text{SLT}} \rangle \approx 5.95$. Some regions of the IGM may have also become transparent to Ly α photons prior to overlap, resulting in windows of transmission inside the Gunn–Peterson trough (one such region may have been seen⁷ in SDSS J1148 + 5251). In the existing examples, the portions of the Universe probed by the lower end of the Gunn–Peterson trough are located several hundred co-moving Mpc away from the background quasar, and are therefore not correlated with the quasar host galaxy. The distribution dP/dz_{SLT} is also independent of the redshift distribution of the quasars. Moreover, lines-of-sight to these quasars are not causally connected at $z \approx 6$ and may be considered independent.

overlap, and hence the reionization epoch, concluded near $z \approx 6$. The most promising observational probe^{8,9} of the reionization epoch is redshifted 21-cm emission from intergalactic H I. Future observations using low-frequency radio arrays (for example, the Low Frequency Array, LOFAR) will allow a direct determination of the topology and duration of the phase of bubble overlap. Here we determine theoretically the expected angular scale and redshift width of the 21-cm fluctuations at the SBO, and show that our determination is consistent with current observational constraints.

We start by quantifying the constraints of light travel time and cosmic variance. First, suppose we have an H II region with

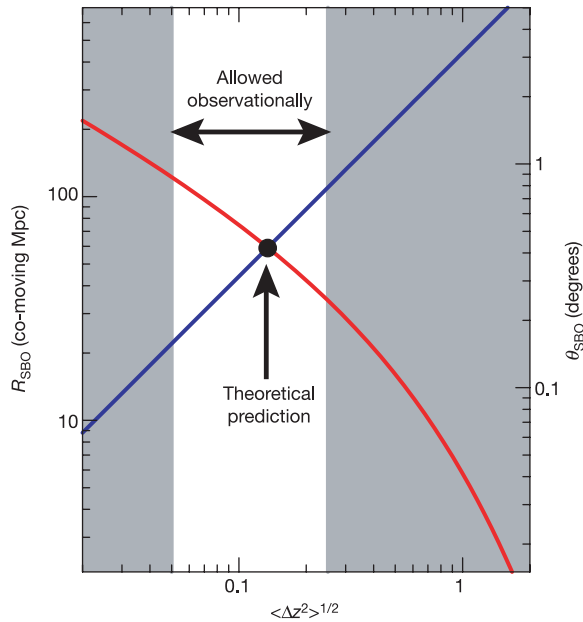


Figure 2 Constraints on the scatter in the SBO redshift and the characteristic size of isolated bubbles at the final overlap stage, R_{SBO} (see Fig. 1). The characteristic size of H II regions grows with time. The SBO is observed for the bubble scale at which the light crossing time (blue line) first becomes smaller than the cosmic scatter in bubble formation times (red line). At $z \approx 6$, the implied scale $R_{\text{SBO}} \approx 60$ co-moving Mpc (or ~ 8.6 physical Mpc), corresponds to a characteristic angular radius of $\theta_{\text{SLT}} \approx 0.4$ degrees on the sky. After bubble overlap, the ionizing intensity grows to a level at which the IGM becomes transparent to Ly α photons. The collapsed fraction required for Ly α transmission within a region of a certain size will be larger than required for its ionization. However, the scatter in equation (2) is not sensitive to the collapsed fraction, and so may be used for both the SBO and SLT. The scatter in the SLT is smaller than the cosmic scatter in the structure formation time on the scale of the mean-free-path for ionizing photons. This mean-free-path must be longer than $R_{\text{SBO}} \approx 60$ Mpc, an inference which is supported by analysis of the Ly α 'forest' at $z \approx 4$ where the mean-free-path is estimated¹⁵ to be ~ 120 co-moving Mpc at the Lyman limit (and longer at higher frequencies). If it is dominated by cosmic variance, then the scatter in the SLT redshift provides a lower limit to the SBO scatter. The three known quasars at $z > 6.1$ have Ly α transmission redshifts of^{4,7} $z_{\text{SLT}} = 5.9, 5.95$ and 5.98 , implying that the scatter in the SBO must be ≥ 0.05 (this scatter may become better known from follow-up spectroscopy of gamma ray burst afterglows at $z > 6$ that might be discovered by the SWIFT satellite^{16,17}). The observed scatter in the SLT redshift is somewhat smaller than the predicted SBO scatter, confirming the expectation that cosmic variance is smaller at the SLT. The scatter in the SBO redshift must also be ≤ 0.25 because the lines-of-sight to the two highest redshift quasars have a redshift of Ly α transparency at $z \approx 6$, but a neutral fraction that is known from the proximity effect¹³ to be substantial at $z \geq 6.2$ – 6.3 . The excluded regions of scatter for the SBO are shown in grey. Throughout this Letter, we adopt the latest values for the cosmological parameters as inferred from the Wilkinson Microwave Anisotropy Probe data¹⁸.

a physical radius $R/(1 + \langle z \rangle)$. The light crossing time of this radius is

$$\langle \Delta z^2 \rangle^{1/2} = \left| \frac{dz}{dt} \right| \frac{R}{c(1 + \langle z \rangle)} \quad (1)$$

where at the high redshifts of interest $(dz/dt) = -(H_0 \sqrt{\Omega_m}) \times (1 + z)^{5/2}$. Here, c is the speed of light, H_0 is the present-day Hubble constant, Ω_m is the present-day matter density parameter, and $\langle z \rangle$ is the mean redshift of the SBO. Note that when discussing this crossing time, we are referring to photons used to probe the ionized bubble (for example, at 21 cm), rather than photons involved in the dynamics of the bubble evolution.

Second, overlap would have occurred at different times in different regions of the IGM owing to the cosmic scatter in the process of structure formation within finite spatial volumes¹. Reionization should be completed within a region of co-moving radius R when the fraction of mass incorporated into collapsed objects in this region attains a certain critical value, corresponding to a threshold number of ionizing photons emitted per baryon. The ionization state of a region is governed by the enclosed ionizing luminosity, by its over-density, and by dense pockets of neutral gas that are self shielding to ionizing radiation. There is an offset¹ δz between the redshift when a region of mean over-density $\bar{\delta}_R$ achieves this critical collapsed fraction, and the redshift $\bar{z}$ when the Universe achieves the same collapsed fraction on average. This offset may be computed¹ from the expression for the collapsed fraction¹⁰ F_{col} within a region of over-density $\bar{\delta}_R$ on a co-moving scale R ;

$$F_{\text{col}}(M_{\text{min}}) = \text{erfc} \left[\frac{\delta_c - \bar{\delta}_R}{\sqrt{2[\sigma_{R_{\text{min}}}^2 - \sigma_R^2]}} \right]$$

which yields

$$\frac{\delta z}{(1 + \bar{z})} = \frac{\bar{\delta}_R}{\delta_c(\bar{z})} - \left[1 - \sqrt{1 - \frac{\sigma_R^2}{\sigma_{R_{\text{min}}}^2}} \right] \quad (2)$$

where $\delta_c(\bar{z}) \propto (1 + \bar{z})$ is the collapse threshold for an over-density at a redshift $\bar{z}$; σ_R and $\sigma_{R_{\text{min}}}$ are the variances in the power spectrum linearly extrapolated to $z = 0$ on co-moving scales corresponding to the region of interest and to the minimum galaxy mass M_{min} , respectively. The offset in the ionization redshift of a region depends on its linear over-density, $\bar{\delta}_R$. As a result, the distribution of offsets, and therefore the scatter in the SBO, may be obtained directly from the power spectrum of primordial inhomogeneities. As can be seen from equation (2), larger regions have a smaller scatter owing to their smaller cosmic variance.

Note that equation (2) is independent of the critical value of the collapsed fraction required for reionization. Moreover, our numerical constraints are very weakly dependent on the minimum galaxy mass, which we choose to have a virial temperature of 10^4 K corresponding to the cooling threshold of primordial atomic gas. The growth of an H II bubble around a cluster of sources requires that the mean-free-path of ionizing photons be of the order of the bubble radius or larger. Since ionizing photons can be absorbed by dense pockets of neutral gas inside the H II region, the necessary increase in the mean-free-path with time implies that the critical collapsed fraction required to ionize a region of size R increases as well. This larger collapsed fraction affects the redshift at which the region becomes ionized, but not the scatter in redshifts from place to place, which is the focus of this Letter. Our results are therefore independent of assumptions about unknown quantities such as the star formation efficiency and the escape fraction of ionizing photons from galaxies, as well as unknown processes of feedback in galaxies and clumping of the IGM.

Figure 2 displays our two fundamental constraints. The light travel time constraint (equation (1)) is shown as the blue line, giving a longer crossing time for a larger bubble size. This contrasts with the constraint of cosmic variance (equation (2)), indicated by the red line, which shows how the scatter in formation times decreases with increasing bubble size. The scatter in the SBO redshift and the corresponding fluctuation scale of the SBO are given by the intersection of these curves. We find that the thickness of the SBO is $\langle \Delta z^2 \rangle^{1/2} \approx 0.13$, and that the bubbles which form the SBO have a characteristic co-moving size of ~ 60 Mpc (equivalent to 8.6 physical Mpc). At $z \approx 6$, this size corresponds to angular scales of $\theta_{\text{SBO}} \approx 0.4$ degrees on the sky. We have also examined a second model for the formation of H II regions, where ionization is reached when the collapsed fraction divided by the density contrast exceeds a critical value. This model allows for the possibility that the increased recombination rate offsets reionization in over-dense regions. The increased range of formation times in this case leads to a slightly larger value for the scale of overlapping H II regions, and we find $\langle \Delta z^2 \rangle^{1/2} \approx 0.2$, $R_{\text{SBO}} \approx 90$ Mpc and $\theta_{\text{SBO}} \approx 0.7$ degrees.

A scatter of ~ 0.15 in the SBO is somewhat larger than the value extracted from existing numerical simulations^{11,12}. The difference is probably due to the limited size of the simulated volumes; although the simulations appropriately describe the reionization process within limited regions of the Universe, they are not sufficiently large to describe the global properties of the overlap phase¹. The scales over which cosmological radiative transfer has been simulated are smaller than the characteristic extent of the SBO, which we find to be $R_{\text{SBO}} \approx 70$ co-moving Mpc.

We can constrain the scatter in the SBO redshift observationally using the spectra of the highest-redshift quasars. Since only a trace amount of neutral hydrogen is needed to absorb Ly α photons, the time where the IGM becomes Ly α -transparent need not coincide with bubble overlap. Following overlap, the IGM was exposed to ionizing sources in all directions and the ionizing intensity rose rapidly. After some time, the ionizing background flux was sufficiently high that the H I fraction fell to a level at which the IGM allowed transmission of resonant Ly α photons. This is shown schematically in Fig. 1. The lower wavelength limit of the Gunn–Peterson trough corresponds to the Ly α wavelength at the redshift when the IGM started to allow transmission of Ly α photons along that particular line-of-sight. In addition to the SBO, we therefore also define the surface of Ly α transmission (hereafter SLT) as the redshift along different lines-of-sight when the diffuse IGM became transparent to Ly α photons.

The scatter in the SLT redshift is an observable which we would like to compare with the scatter in the SBO redshift. The variance of the density field on large scales results in the biased clustering of sources¹. H II regions grow in size around these clusters of sources. In order for the ionizing photons produced by a cluster to advance the walls of the ionized bubble around it, the mean-free-path of these photons must be of the order of the bubble size or larger. After bubble overlap, the ionizing intensity at any point grows until the ionizing photons have time to travel across the scale of the new mean-free-path, which represents the horizon out to which ionizing sources are visible. Since the mean-free-path is larger than R_{SBO} , the ionizing intensity at the SLT averages the cosmic scatter over a larger volume than at the SBO. This constraint implies that the cosmic variance in the SLT redshift must be smaller than the scatter in the SBO redshift. However, it is possible that opacity from small-scale structure contributes additional scatter to the SLT redshift.

If cosmic variance dominates the observed scatter in the SLT redshift, then, on the basis of the spectra of the three $z > 6.1$ quasars^{4,7}, we would expect the scatter in the SBO redshift to satisfy $\langle \Delta z^2 \rangle_{\text{obs}}^{1/2} \gtrsim 0.05$. In addition, analysis of the proximity effect for the size of the H II regions around the two highest-redshift quasars^{13,14}

implies a neutral fraction that is of the order of unity (that is, pre-overlap) at $z \approx 6.2$ – 6.3 , while the transmission of Ly α photons at $z \lesssim 6$ implies that overlap must have completed by that time. This restricts the scatter in the SBO to be $\langle \Delta z^2 \rangle_{\text{obs}}^{1/2} \lesssim 0.25$. The constraints on values for the scatter in the SBO redshift are shaded grey in Fig. 2. It is reassuring that the theoretical prediction for the SBO scatter of $\langle \Delta z^2 \rangle^{1/2} \approx 0.15$, with a characteristic scale of ~ 70 co-moving Mpc, is bounded by these constraints.

The presence of significantly neutral IGM just beyond the redshift of overlap^{13,14} is encouraging for upcoming 21-cm studies of the reionization epoch, as it results in emission near an observed frequency of 200 MHz where the signal is most readily detectable. Future observations of redshifted 21-cm line emission at $6 \lesssim z \lesssim 6.5$ with instruments such as LOFAR will be able to map the three-dimensional distribution of H I at the end of reionization. The intergalactic H II regions will imprint a ‘knee’ in the power spectrum of the 21-cm anisotropies on a characteristic angular scale corresponding to a typical isolated H II region⁸. Our results suggest that this characteristic angular scale is large at the end of reionization, $\theta_{\text{SBO}} \approx 0.5$ degrees, motivating the construction of compact low-frequency arrays. An SBO thickness of $\langle \Delta z^2 \rangle^{1/2} \approx 0.15$ suggests a minimum frequency bandwidth of ~ 8 MHz for experiments aiming to detect anisotropies in 21-cm emission just prior to overlap. These results will help guide the design of the next generation of low-frequency radio observatories in the search for 21-cm emission at the end of the reionization epoch. $\square$

Received 23 July; accepted 14 September 2004; doi:10.1038/nature03033.

1. Barkana, R. & Loeb, A. Unusually large fluctuations in the statistics of galaxy formation at high redshift. *Astrophys. J.* **609**, 474–481 (2004).
2. Furlanetto, S. R., Zaldarriaga, M. & Hernquist, L. The growth of H II regions during reionization. *Astrophys. J.* (2004) (submitted); preprint astro-ph/0403697 at (<http://arxiv.org>).
3. Gunn, J. E. & Peterson, B. A. On the density of neutral hydrogen in intergalactic space. *Astrophys. J.* **142**, 1633–1641 (1965).
4. Fan, X. *et al.* A survey of $z > 5.7$ quasars in the Sloan Digital Sky Survey III: discovery of five additional quasars. *Astron. J.* (2004) (in the press); preprint astro-ph/0405138 at (<http://arxiv.org>).
5. Fan, X. *et al.* Evolution of the ionizing background and the epoch of reionization from the spectra of $z \sim 6$ quasars. *Astron. J.* **123**, 1247–1257 (2002).
6. Pentericci, L. *et al.* VLT optical and near-infrared observations of the $z = 6.28$ quasar SDSS J1030 + 0524. *Astron. J.* **123**, 2151–2158 (2002).
7. White, R. L., Becker, R. H., Fan, X. & Strauss, M. A. Probing the ionization state of the universe at $z > 6$. *Astron. J.* **126**, 1–14 (2003).
8. Zaldarriaga, M., Furlanetto, S. R. & Hernquist, L. 21 centimeter fluctuations from cosmic gas at high redshifts. *Astrophys. J.* (2003) (submitted); preprint astro-ph/0311514 at (<http://arxiv.org>).
9. Morales, M. F. & Hewitt, J. Toward epoch of reionization measurements with wide-field radio observations. *Astrophys. J.* (2003) (in the press); preprint astro-ph/0312437 at (<http://arxiv.org>).
10. Bond, J. R., Cole, S., Efstathiou, G. & Kaiser, N. Excursion set mass functions for hierarchical Gaussian fluctuations. *Astrophys. J.* **379**, 440–460 (1991).
11. Gnedin, N. Y. Cosmological reionization by stellar sources. *Astrophys. J.* **535**, 530–554 (2000).
12. Yoshida, N., Sokasian, A., Hernquist, L. & Springel, V. Early structure formation and reionization in a cosmological model with a running primordial power spectrum. *Astrophys. J.* **598**, 73–85 (2003).
13. Wyithe, J. S. B. & Loeb, A. A large neutral fraction of cosmic hydrogen a billion years after the Big Bang. *Nature* **427**, 815–817 (2004).
14. Mesinger, A. & Haiman, Z. Evidence for a cosmological Stromgren surface and for significant neutral hydrogen surrounding the quasar SDSS J1030 + 0524. *Astrophys. J. Lett.* (2004) (in the press); preprint astro-ph/0406188 at (<http://arxiv.org>).
15. Miralda-Escudé, J. On the evolution of the ionizing emissivity of galaxies and quasars required by the hydrogen reionization. *Astrophys. J.* **597**, 66–73 (2003).
16. Barkana, R. & Loeb, A. GRBs versus quasars: Lyman- α signatures of reionization versus cosmological infall. *Astrophys. J.* **601**, 64–77 (2004).
17. Bromm, V. & Loeb, A. The expected redshift distribution of gamma-ray burst. *Astrophys. J.* **575**, 111–116 (2002).
18. Spergel, D. N. *et al.* First-year Wilkinson microwave anisotropy probe (WMAP) observations: determination of cosmological parameters. *Astrophys. J. Suppl.* **148**, 175–194 (2003).

Acknowledgements This work was supported in part by ARC, NSF and NASA.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to J.S.B.W. (swyithe@isis.ph.unimelb.edu.au) or A.L. (aloeb@cfa.harvard.edu).

Strong coupling in a single quantum dot–semiconductor microcavity system

J. P. Reithmaier¹, G. Şek^{1*}, A. Löffler¹, C. Hofmann¹, S. Kuhn¹, S. Reitzenstein¹, L. V. Keldysh², V. D. Kulakovskii³, T. L. Reinecke⁴ & A. Forchel¹

¹Technische Physik, Universität Würzburg, Am Hubland, D-97074 Würzburg, Germany

²Lebedev Physical Institute, Russian Academy of Science, 119991 Moscow, Russia

³Institute for Solid State Physics, Russian Academy of Science, 142432 Chernogolovka, Russia

⁴Naval Research Laboratory, Washington, DC 20375, USA

* Permanent address: Institute of Physics, Wrocław University of Technology, 50-370 Wrocław, Poland

Cavity quantum electrodynamics, a central research field in optics and solid-state physics^{1–3}, addresses properties of atom-like emitters in cavities and can be divided into a weak and a strong coupling regime. For weak coupling, the spontaneous emission can be enhanced or reduced compared with its vacuum level by tuning discrete cavity modes in and out of resonance with the emitter^{2,4–13}. However, the most striking change of emission properties occurs when the conditions for strong coupling are fulfilled. In this case there is a change from the usual irreversible spontaneous emission to a reversible exchange of energy between the emitter and the cavity mode. This coherent coupling may provide a basis for future applications in quantum information processing or schemes for coherent control. Until now, strong coupling of individual two-level systems has been observed only for atoms in large cavities^{14–17}. Here we report the observation of strong coupling of a single two-level solid-state system with a photon, as realized by a single quantum dot in a semiconductor microcavity. The strong coupling is manifest in photoluminescence data that display anti-crossings between the quantum dot exciton and cavity-mode dispersion relations, characterized by a vacuum Rabi splitting of about 140 µeV.

Strong coupling occurs when the emitter–photon interaction becomes larger than the combined atomic dipole decay rate and the cavity field decay rate. Then the irreversible spontaneous emission process of the emitter is replaced by a coherent periodic energy exchange between the emitter and the photon in the form of Rabi oscillations for timescales shorter than the inverse cavity field decay rate. In spectroscopic experiments this energy exchange results in anti-crossings between the atom-like emitter and cavity-mode dispersion relations and is characterized by the vacuum Rabi splitting. Solid state implementations, which would be highly desirable for future applications, for example in the area of quantum information processing^{18–22}, have not yet been achieved.

Semiconductor heterostructures are the best candidates for the observation of strong coupling in solids, because they permit the realization of solid state cavities in which atom-like emitters in the form of quantum dots (QDs) can be embedded. In these QD semiconductor cavities, strong and weak interaction can occur between the QD exciton (X) and discretized cavity (C) modes at a resonance ($E_X = E_C = E_0$). In a picture of coupled oscillators the energies of the interacting modes at resonance are^{23,24}

$$E_{1,2} = E_0 - i(\gamma_C + \gamma_X)/4 \pm [g^2 - (\gamma_C - \gamma_X)^2/16]^{1/2} \quad (1)$$

where $\gamma_{C,X}$ is the full width at half maximum (FWHM) of the cavity and exciton modes, respectively, and g is the exciton–photon coupling parameter. Strong coupling requires $g^2 > (\gamma_C - \gamma_X)^2/16$. This regime is characterized by a gap corresponding to the

vacuum Rabi splitting between the two energies $E_{1,2}$ on resonance. In contrast, for $g^2 \leq (\gamma_C - \gamma_X)^2/16$ the real parts of the energies $E_{1,2}$ are degenerate. This corresponds to the weak coupling case characterized, for example, by an enhancement of the spontaneous emission on resonance by the Purcell effect^{2–3,5–7,25}.

The exciton–photon coupling parameter g is given by the scalar product of the transition matrix element of the QD exciton dipole moment with the local value of the electric field at the dot position. Assuming that the QD is located at the antinode of the electromagnetic field of the cavity mode, the coupling constant is related to the oscillator strength f and the mode volume V_m by²³

$$g = (\pi e^2 f)^{1/2} / (4\pi \epsilon_r \epsilon_0 m_0 V_m)^{1/2} \quad (2)$$

where ϵ_r and ϵ_0 are the dielectric constants of cavity material and vacuum, respectively, and m_0 is the free electron mass. Equation (2) shows that g depends on the QD exciton oscillator strength f and on the mode volume V_m as $(f/V_m)^{1/2}$.

γ_C is related to the quality factor of the cavity $Q = E_C/\gamma_C$. The intrinsic value of the FWHM of the QD exciton γ_X is on the order of a few µeV (ref. 26), which is much smaller than γ_C (~100 µeV or larger in the present cavities; see Supplementary Information). The criterion for strong coupling can therefore be approximated by $g > \gamma_C/4$. This corresponds to the need to maximize the product $(f/V_m)^{1/2}Q$ so as to overcome the threshold for strong coupling. That is, one must realize QD excitons with large oscillator strength in high- Q cavities with small mode volumes. In addition, because the coupling is given by the electromagnetic field strength at the exciton position, the QD should be located at the antinode of the field in all three dimensions (see Supplementary Information).

We have investigated strong coupling effects of a single QD with discrete cavity modes in a semiconductor micropillar cavity. Details of the structure are given in the Supplementary Information. The epitaxial structures provide strong optical confinement in the growth direction by Bragg reflector stacks composed of 20 and 23 periods of GaAs and AlAs layers in the top and bottom mirrors surrounding a GaAs cavity. At the centre of the GaAs cavity at the antinode of the fundamental mode in growth direction an InGaAs QD layer is embedded.

As discussed above, one has to maximize the oscillator strength of the QD exciton as well as the cavity quality factor Q at small mode volume V_m so as to reach strong coupling for single QDs. The oscillator strength of the QD excitons increases with the dot size. It has therefore been suggested²³ that large ('natural') QDs be used,

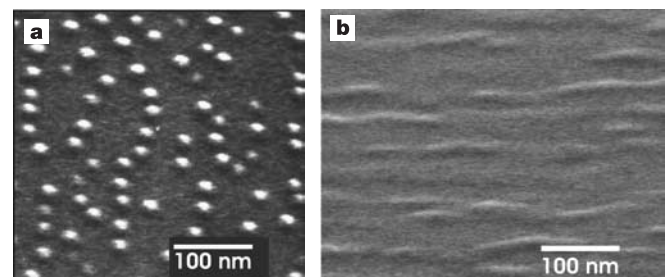


Figure 1 Scanning electron micrographs of InGaAs QDs with different In contents before overgrowth. The samples have been tilted with respect to the electron beam direction. The magnification along the vertical axis is therefore about one-third of that along the horizontal direction. **a**, For an indium content of 60%, typical self-assembled dots are formed owing to the large lattice mismatch to the GaAs substrate accompanied by a large resulting strain. These QDs have a characteristic diameter of 15–20 nm. **b**, Much larger dots are formed for an In content of 30%, as used in the present micropillars. These dots are usually elongated with lengths on the order of 100 nm or larger and widths of about 30 nm. The larger size gives rise to an increase in the excitonic oscillator strength compared with the smaller self-assembled dots.

which can be obtained by growth fluctuations of quantum wells instead of small self-assembled dots. For the present work we have realized $\text{In}_{0.3}\text{Ga}_{0.7}\text{As}$ natural dots. As shown by the scanning electron micrographs in Fig. 1, the use of 30% In results in large asymmetric dots with typical lengths of 100 nm and widths of about 30 nm. Conventional self-assembled dots shown in Fig. 1 for an In content of 60% are characterized by a more symmetric shape and diameters of 15–20 nm. By using natural dots with 30% In content the dot area can therefore be increased by about an order of magnitude, which should result in a significant enhancement of the oscillator strength.

Photoluminescence spectroscopy on the as-grown sample shows a sharp cavity emission with Q factors of about 12,000. By removing the upper Bragg reflector the emission of the QD layer can be studied without cavity effects. At low temperature (4 K to about 50 K) the QD layer exhibits a broad photoluminescence spectrum in the energy range 1.33–1.35 eV because of inhomogeneous broadening due to fluctuations in QD size.

To obtain high- Q cavities with small mode volumes as required for the observation of strong coupling effects, we have fabricated micropillars of various sizes from the epitaxial samples. Micropillars with circular cross-sections (diameters from 0.5 to 4 μm) have been processed by electron-beam lithography and reactive ion-etching in an inductively coupled Ar/ Cl_2 plasma. Figure 2a shows such a micropillar with a diameter of 800 nm and a height of about 6 μm . The combination of a high- Q epitaxial cavity with optimized micropillar processing allows us to realize micropillars with Q factors of 8,000–9,000 for 2 μm diameter, 7,000–9,000 for 1.5 μm diameter and 2,000–4,000 for 1.0 μm diameter.

For cylindrical micropillars with constant cavity height and different pillar diameters the need to maximize $(f/V_{\text{m}})^{1/2}Q$ for the observation of strong coupling corresponds to the maximization of Q/d_{C} at a given oscillator strength f . Here d_{C} denotes the diameter of the pillars. For cavities with diameters of 2, 1.5 and 1.0 μm and Q factors at the centre of the ranges given above, we obtain values of 4,250, 5,300 and 3,000 μm^{-1} for Q/d_{C} , respectively. This implies

that the 1.5- μm diameter micropillars offer the best conditions for the observation of strong coupling.

For single-dot photoluminescence studies, samples with micropillars 1.5 μm in diameter were mounted in a variable-temperature cryostat of a microphotoluminescence set-up. Luminescence of single dots in individual pillars is spectrally resolved by a 1-m double spectrometer equipped with a nitrogen-cooled charge-coupled-device camera. The spectral resolution of the system amounts to about 50 μeV . For optical excitation the beam of a frequency-doubled continuous-wave Nd:YAG laser at 532 nm was focused into a spot 3 μm in diameter centred at the pillar by a microscope objective, which was also used to collect the emission of the sample.

Figure 2b shows a photoluminescence spectrum of one of the 1.5- μm -diameter pillars. The spectrum is dominated by the emission of the cavity mode located at 1.316 eV. The optical mode has a FWHM of 0.15 meV, corresponding to a Q factor of 8,800. In the high- Q cavities investigated here, the emission of the individual dots results in lines of lower intensity and smaller FWHM. The observed FWHM of the single-dot lines located outside the interaction range with the cavity mode indicates the spectral resolution of the experiment.

To investigate the coupling between the cavity mode and individual QDs, the energies of both must be tuned through each other. For tuning through resonance, the different temperature variations of the bandgap and of the refractive index n are used. The temperature dependence of the emission energy of the QD excitons (typically -0.04 meV K^{-1} at 25 K) is governed by the temperature dependence of the bandgap, whereas the temperature dependence of the cavity modes (about $-0.005 \text{ meV K}^{-1}$ at 25 K) is determined by the much weaker temperature dependence of the refractive index. By selecting micropillars with single-QD exciton emissions near to the energy of the cavity resonance it is possible to tune the excitations in and out of resonance.

Figure 3 shows spectra taken from a microcavity 1.5 μm in diameter at different temperatures between 5 and 30 K. At 5 K the emission consists of the cavity mode C centred at about 1.32335 eV and a QD exciton X emission at slightly higher energy (1.32365 eV). By using lineshape fits of the 5 K spectrum with two Lorentzians, we find that the emission intensity of the dot exciton is smaller by about one order of magnitude than that of the cavity mode. The FWHM of the cavity mode (180 μeV for low-excitation conditions) is found to be larger by about 2.5-fold than the FWHM of the dot. For

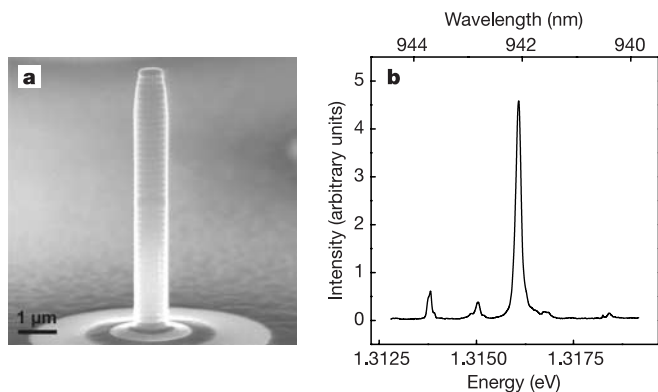


Figure 2 Scanning electron micrograph and photoluminescence of processed microcavity pillars. **a**, Scanning electron micrograph of a pillar with a diameter of about 0.8 μm . By a combination of electron-beam lithography and reactive dry etching, micropillars with close to vertical and defect-free sidewalls are obtained, as required for high- Q cavities. **b**, Microphotoluminescence spectrum for a 1.5- μm diameter microcavity ($Q = 8,800$) with $\text{In}_{0.3}\text{Ga}_{0.7}\text{As}$ QDs recorded at 5 K. A few sharp (spectral-resolution-limited linewidths) peaks related to the emission from QDs and a strong and wider peak connected to the fundamental cavity mode are observed. Even without a dot on resonance, there is cavity emission due to population through, for example, phonon-assisted scattering from the dot ensemble in the pillar. The cavity peak is particularly pronounced owing to the high Q factors of the present structures. QD lines that show, at low temperature, energies slightly higher than that of the cavity mode can be temperature tuned through the cavity resonance.

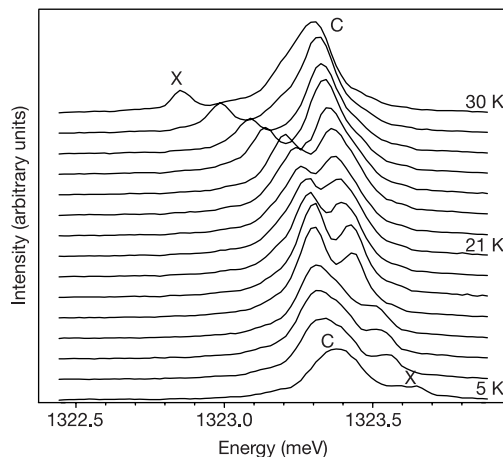


Figure 3 Temperature dependence of photoluminescence spectra for a 1.5- μm microcavity ($Q = 7,350$) showing the tuning of a single QD exciton through resonance with the cavity mode. A clear anti-crossing is observed owing to strong coupling between the QD exciton and the microcavity photon mode.

increasing temperature up to 30 K the emission consists of two distinct features. However, at 30 K the components of the emission have exchanged their properties: now the low-energy line has an emission intensity and FWHM similar to that of the line located at 5 K at higher energy and therefore has to be assigned to the QD exciton. Simultaneously, the high-energy line displays an emission intensity and FWHM that correspond to the values for the cavity mode at 5 K. Over the entire temperature range the energies of the two contributions to the spectrum are well separated and avoid crossing each other. All the above findings are clear indications for an anti-crossing of the single QD exciton and cavity-mode dispersions due to strong coupling between those modes.

Figure 4 shows experimentally observed variations of the energies, FWHMs and intensities of the coupled modes for varying temperature for two 1.5 μm micropillars exhibiting strong and weak coupling, respectively. The data are plotted against the change of the energies of a different, non-interacting QD in the same cavity that is separated by about 5 meV from the cavity-mode position. The origin of this energy scale has been placed at the resonance.

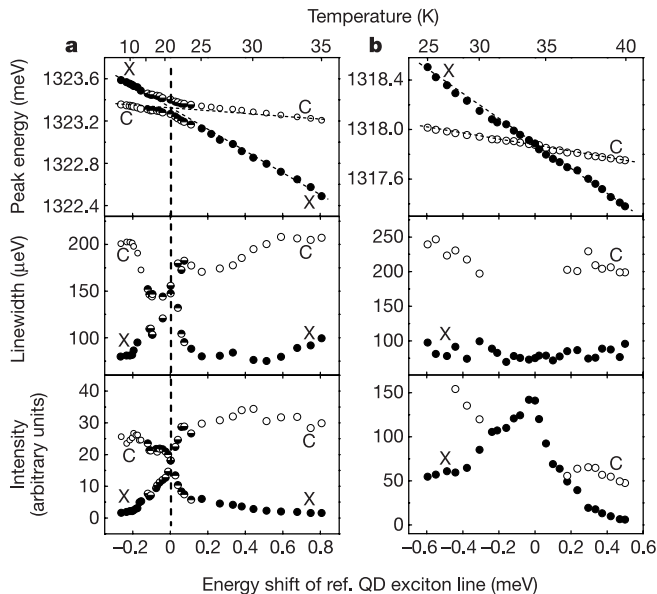


Figure 4 Dependences of photoluminescence peak energies, linewidths and integrated intensities on the temperature-generated energy shift of a reference QD for strong and weak coupling. The reference QD-related peak is located about 5 meV above the cavity mode and does not interact with the cavity (bottom scale). At the top is the temperature axis (nonlinear). All data have been obtained by fits of lorentzian lineshapes to the experimental spectra. **a**, Data for a single dot–cavity system in strong coupling. C (open dots), cavity-mode data; X (filled dots), QD exciton data; partly filled dots, data on or near resonance, where the strong coupling results in coupled modes. Top, anti-crossing of QD exciton and cavity mode. Centre, variation of exciton, cavity and coupled-mode FWHM linewidths during anti-crossing. Bottom, exchange of intensities of single-dot exciton and cavity mode (integrated over the respective lorentzian used for the fits) resulting from anti-crossing. **b**, Experimental results for a dot–cavity system in the weak coupling regime. Top, crossing of cavity (C, open dots) and single QD exciton (X, filled dots) dispersions. Centre, variation of single-dot exciton FWHM linewidths during crossing of the dispersions (open circles). Values for the cavity mode are shown only above and below the resonance, because a reliable determination of the cavity FWHM on resonance is prevented by the small cavity contribution. Bottom, variation of integrated single-dot exciton and cavity-mode intensities. The dot intensity on resonance shows a clear peak due to the Purcell effect. In comparison with those in **a**, the experiments shown in **b** were performed at elevated temperatures between 25 and 40 K. Because of increasing thermal ionization, the plot of intensity against detuning shows a decrease in QD emission intensity superimposed on the Purcell effect peak.

As plotted in the top panel of Fig. 4a for a strongly coupled system, the energies of the higher-lying mode show a fairly strong variation before reaching the resonance (below about 20 K), which changes to a much weaker dependence. Simultaneously the low-energy mode, which displays only a weak temperature variation at low temperature, changes to a stronger temperature dependence above 20 K. The top panel of Fig. 4b shows experimental data for the energy variation of both modes for a micropillar with dot exciton and cavity mode in a weak coupling situation during temperature tuning. In striking contrast to the strong coupling case shown in Fig. 4a, the exciton dispersion with its steep characteristic temperature dependence crosses the weakly temperature-dependent cavity mode without any indication of a gap. By comparing the temperature dependences of the strongly coupled modes in Fig. 4a with the temperature variations of weakly interacting QD excitons and cavity modes in Fig. 4b, we observe that weakly interacting excitons have similar temperature dependences to the steep sections of the dispersions of the coupled modes. The less temperature dependent sections of the coupled mode dispersions correspond to the temperature variations of uncoupled cavity modes. For the upper branch the anti-crossing results in a change from an exciton-like dispersion to a cavity-mode-like one, whereas the lower branch changes from a cavity-mode dispersion to that of an exciton with increasing temperature.

As shown in the top panel of Fig. 4a for the case of strong coupling, the energy separation of the coupled modes on resonance amounts to about 140 μeV . This corresponds to the vacuum Rabi splitting of this coupled single two-level emitter (QD exciton) solid state cavity system. Using equation (1) and the experimental values for the FWHM of the cavity mode out of resonance and the observed splitting, we estimate a value of $g \approx 0.08$ meV for the coupling constant. The threshold for the onset of strong coupling, which is given by a vanishing square root term in equation (1), corresponds to $g = 0.045$ meV.

By numerical calculations we obtain mode volumes of about $0.3 \mu\text{m}^3$ for the present micropillar cavities. Using this value and equation (2) we estimate the effective oscillator strength to about 50. This value is about fourfold higher than those reported for self-assembled InAs QDs²³. However, it agrees with values observed for natural QDs in the GaAs/AlGaAs system²⁷. The use of dots with large oscillator strengths is crucial for the observation of strong coupling: from equation (1), for the present Q/d_C values and small self-assembled InAs dots with an oscillator strength of about 10, the QD–cavity system would still be in the weak coupling regime.

In the middle and bottom panels of Fig. 4 the variations of the linewidths (FWHM) and intensities of the cavity–dot system for the cases of strong and weak coupling are compared with each other. For strong coupling we observe an exchange of linewidths and intensities at the anti-crossing (Fig. 4a). As a result of the anti-crossing, both contributions to the emission can be analysed reliably over the entire temperature range. Because of the absence of a gap in the weak coupling regime, the analysis of the individual FWHMs and intensities of the two components is more complex than for the strong coupling case (see Supplementary Fig. 2). We therefore include only those features in Fig. 4b that can be evaluated reliably. These are the emission intensity and FWHM of the QD exciton. For the weak coupling, the emission intensity of the dot increases strongly on resonance owing to the Purcell effect^{4–9,25}. Within experimental accuracy we find no broadening for the dot line due to interaction with the cavity mode.

The present results for strong coupling are due to the coherent interaction of a single QD emitter with a single photon. As described in the Supplementary Information there is no indication for the contribution of two dots when the dots are far from resonance. Because the oscillator strength derived from our experiment also agrees with values for excitons in single QDs, we conclude that there is only one dot participating in the strong

coupling. The excitation power used in the strong coupling experiments of $2\mu\text{W}$ corresponds to 1.5×10^9 photons s^{-1} in the cavity averaged over time; that is, interactions of the single QD with multiple photons can be neglected.

Our experiments demonstrate that long-sought solid state implementations of the strongly coupled cavity-mode-two-level-emitter systems are feasible by using single QDs in high-Q cavities with small mode volumes. With further improvements, for example using higher-Q cavities or QDs placed at the in-plane mode centre, these systems have the potential for wide application ranging from nonlinear optics²⁸ to quantum information processing^{18–22}. □

Received 11 June; accepted 26 August 2004; doi:10.1038/nature02969.

1. Yamamoto, Y. & Slusher, R. E. Optical processes in microcavities. *Physics Today* **46**, 66–73 (1993).
2. Gerard, J. M. & Gayral, B. InAs quantum dots: artificial atoms for solid-state cavity-quantum electrodynamics. *Physica E* **9**, 131–139 (2001).
3. Vahala, K. J. Optical microcavities. *Nature* **424**, 839–846 (2003).
4. Kleppner, D. Inhibited spontaneous emission. *Phys. Rev. Lett.* **47**, 233–236 (1981).
5. Goy, P., Raimond, J. M., Cross, M. M. & Haroche, S. Observation of cavity-enhanced single-atom spontaneous emission. *Phys. Rev. Lett.* **50**, 1903–1906 (1983).
6. Gabrielse, G. & Dehmelt, H. Observation of inhibited spontaneous emission. *Phys. Rev. Lett.* **55**, 67–70 (1985).
7. Hulet, R. G., Hilfer, E. S. & Kleppner, D. Inhibited spontaneous emission by a Rydberg atom. *Phys. Rev. Lett.* **55**, 2137–2140 (1985).
8. Gerard, J. M. *et al.* Enhanced spontaneous emission by quantum boxes in a monolithic optical microcavity. *Phys. Rev. Lett.* **81**, 1110–1113 (1998).
9. Bayer, M. *et al.* Inhibition and enhancement of the spontaneous emission of quantum dots in structured microresonators. *Phys. Rev. Lett.* **86**, 3168–3171 (2001).
10. Solomon, G. S., Pelton, M. & Yamamoto, Y. Single-mode spontaneous emission from a single quantum dot in a three-dimensional microcavity. *Phys. Rev. Lett.* **86**, 3903–3906 (2001).
11. Pelton, M. *et al.* Efficient source of single photons: a single quantum dot in a micropost microcavity. *Phys. Rev. Lett.* **89**, 233602–1–4 (2002).
12. Santori, C., Fattal, D., Vuckovic, J., Solomon, G. S. & Yamamoto, Y. Indistinguishable photons from a single-photon device. *Nature* **419**, 594–597 (2002).
13. Michler, P. *et al.* A quantum dot single-photon turnstile device. *Science* **290**, 2282–2285 (2000).
14. Hood, C. J., Chapman, M. S., Lynn, T. W. & Kimble, H. J. Real-time cavity QED with single atoms. *Phys. Rev. Lett.* **80**, 4157–4160 (1998).
15. Mabuchi, H. & Doherety, A. C. Cavity quantum electrodynamics: coherence in context. *Science* **298**, 1372–1377 (2002).
16. McKeever, J. *et al.* Experimental realization of one-atom laser in the regime of strong coupling. *Nature* **425**, 268–271 (2003).
17. McKeever, J. *et al.* State-insensitive cooling and trapping of single atoms in an optical cavity. *Phys. Rev. Lett.* **90**, 133602–1–4 (2003).
18. Monroe, C. Quantum information processing with atoms and photons. *Nature* **416**, 238–246 (2002).
19. Imamoglu, A. *et al.* Quantum information processing using quantum dot spins and cavity QED. *Phys. Rev. Lett.* **83**, 4204–4207 (1999).
20. Stievater, T. H. *et al.* Rabi oscillations of excitons in single quantum dots. *Phys. Rev. Lett.* **87**, 133603–1–4 (2001).
21. Li, X. Q. & Yan, Y. J. Quantum computation with coupled quantum dots in optical microcavities. *Phys. Rev. B* **65**, 205301–1–5 (2002).
22. Kiraz, A., Ataturk, M. & Imamoglu, A. Quantum-dot single-photon sources: Prospects for applications in linear optics quantum-information processing. *Phys. Rev. A* **69**, 032305 (2004).
23. Andreani, L., Panzarini, G. & Gerard, J. M. Strong-coupling regime for quantum boxes in pillar microcavities: Theory. *Phys. Rev. B* **60**, 13276–13279 (1999).
24. Rudin, S. & Reinecke, T. L. Oscillator model for vacuum Rabi splitting in microcavities. *Phys. Rev. B* **59**, 10227–10233 (1999).
25. Purcell, E. M. Spontaneous emission probabilities at radio frequencies. *Phys. Rev.* **69**, 681 (1946).
26. Bayer, M. & Forchel, A. Temperature dependence of the exciton homogeneous linewidths in $\text{In}_{0.6}\text{Ga}_{0.4}\text{As}/\text{GaAs}$ self-assembled quantum dots. *Phys. Rev. B* **65**, 041308–1–4 (R) (2002).
27. Guest, J. R. *et al.* Measurement of optical absorption by a single quantum dot exciton. *Phys. Rev. B* **65**, 241310–1–4 (2002).
28. Shimizu, Y. *et al.* Control of light pulse propagation with only a few cold atoms in a high finesse microcavity. *Phys. Rev. Lett.* **89**, 233001–1–4 (2002).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements Partial financial support of this work by the DARPA QuIST program, the Deutsche Forschungsgemeinschaft via Research Group Quantum Optics in Semiconductor Nanostructures, the Office of Naval Research and the ONR Nanoscale Electronics Program, INTAS and the State of Bavaria is acknowledged.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to A.F. (alfred.forchel@physik.uni-wuerzburg.de).

Vacuum Rabi splitting with a single quantum dot in a photonic crystal nanocavity

T. Yoshie¹, A. Scherer¹, J. Hendrickson², G. Khitrova², H. M. Gibbs², G. Rupper², C. Ell², O. B. Shchekin³ & D. G. Deppe³

¹Electrical Engineering, California Institute of Technology, Pasadena, California 91125, USA

²Optical Sciences Center, The University of Arizona, Tucson, Arizona 85721, USA

³Microelectronics Research Center, Department of Electrical and Computer Engineering, The University of Texas at Austin, Austin, Texas 78712, USA

Cavity quantum electrodynamics (QED) systems allow the study of a variety of fundamental quantum-optics phenomena, such as entanglement, quantum decoherence and the quantum–classical boundary^{1–9}. Such systems also provide test beds for quantum information science. Nearly all strongly coupled cavity QED experiments have used a single atom in a high-quality-factor (high-Q) cavity. Here we report the experimental realization of a strongly coupled system in the solid state: a single quantum dot embedded in the spacer of a nanocavity, showing vacuum-field Rabi splitting exceeding the decoherence linewidths of both the nanocavity and the quantum dot. This requires a small-volume cavity and an atomic-like two-level system^{5,10}. The photonic crystal¹¹ slab nanocavity—which traps photons when a defect is introduced inside the two-dimensional photonic bandgap by leaving out one or more holes¹²—has both high Q and small modal volume V, as required for strong light–matter interactions¹³. The quantum dot has two discrete energy levels with a transition dipole moment much larger than that of an atom^{14–16}, and it is fixed in the nanocavity during growth.

The study of vacuum Rabi splitting has been an exciting subfield of atomic physics since its first observation with many atoms in the early 1980s; see ref. 1 for a history of the field. After a decade of gradually improving the Q of the cavity and decreasing its volume, vacuum Rabi splitting was seen with a single atom. This opened exciting opportunities for the field of atomic cavity QED, and many experiments followed^{1–7}. For such a truly quantum system, the optical properties are changed by the addition of a single photon or single atom, and the quantum–classical boundary can be studied^{2–4}. But because atoms can move and even escape, their coupling is time-dependent; clearly, the next goal was to localize a cold atom inside the cavity using atomic traps⁶.

In the field of semiconductors, 12 years elapsed between seeing non-perturbative normal mode coupling¹⁷, analogous to many-atom vacuum Rabi splitting¹⁰, and the observation of strong coupling between a single quantum dot (SQD) and a small-volume crystal nanocavity. This advance, which produced opportunities for truly quantum-optics cavity QED experiments in semiconductors, owes much to the extensive studies of (and improvements in) SQDs and monolithic cavities. The semiconductor approximation to a two-level system is a SQD, a small semiconductor crystal confined in three dimensions by a higher-bandgap material^{14–16}. The sharp emission lines observed from submicrometre collection spots were shown to arise from transitions between discrete energy levels of the quantum dot (QD) depending upon size and shape^{18,19}. Coherent transient experiments were performed on these atom-like transitions^{16,20}, and their spontaneous emission was enhanced^{21–23} and inhibited²² by the Purcell effect within microcavities²⁴ of higher and higher Q. The transitions of a SQD can be separated enough for the lowest transition to exhibit anti-bunching, and cavity enhanced spontaneous emission can lead to one photon on demand into a desired mode^{21,25}.

The condition for strong coupling is more demanding on Q : the vacuum Rabi splitting, $2g$, due to a SQD must exceed the mean of the decay rates of the cavity, κ , and the dot, γ . The coupling strength, g , is given by $\mu E_{\text{vac}}/\hbar$, where the vacuum field satisfies $n^2 \epsilon_0 |E_{\text{vac}}|^2 V = \hbar \nu/2$. Here $n \approx 3.4$ is the semiconductor refractive index, ϵ_0 is the permittivity of vacuum, V is the mode volume, and ν is the frequency of the transition of the quantum dot with dipole moment μ . For small-length cavities, $\kappa (= \nu/Q)$ usually exceeds γ ; then, since $g/\kappa \propto Q/\sqrt{V}$, the challenge has been to fabricate a high- Q cavity while maintaining a very small V . Recently, a breakthrough in design by Noda's group¹³ resulted in silicon photonic crystal nanocavities with $Q = 45,000$ and $V = 0.07 \mu\text{m}^3$. Clearly, the cavity with the smallest V (while maintaining high Q) yields strong coupling with a smaller dipole, that is, the dot is more quantum.

Our photonic crystal nanocavity follows the design¹³ of Noda's group, but the silicon is replaced by GaAs for growth of quantum dots. The three-dimensional-mode in-plane confinement is obtained by fabricating a two-dimensional triangular lattice photonic crystal slab with three holes missing to form a spacer (Fig. 1a, b). The vertical confinement, achieved by total internal reflection at the slab semiconductor–air interfaces, is imperfect, in that light with

small in-plane wavevectors can leak out of the top and bottom. Noda's one-dimensional model showed that the key to reducing this loss is to shift out slightly the holes at the ends of the spacer: "the light has to be confined gently in order to confine it strongly"¹³. In other words, when the field envelope function is stopped abruptly, its Fourier transform has a larger overlap with small in-plane wavevectors that leak out; terminating it gently reduces that loss¹³.

The sample, grown on a (001) GaAs substrate by molecular beam epitaxy, has a single layer of InAs quantum dots in the centre of the slab²⁶ (Fig. 2a). A large array of nanocavities ($\sim 30,000$ in clusters of 30 with a density of $5,560 \text{ cavities mm}^{-2}$) is fabricated with crystal parameters changed systematically. The missing-holes spacer is surrounded by 14 periods of air holes for good in-plane optical confinement. The parameters of the photonic crystal are controlled lithographically²⁶: $a = 300 \text{ nm}$, $r = 0.27a$, $s = 0.20a$, and slab thickness $d = 0.90a$ (see Fig. 1b for definitions of a , r and s).

Computations of the field strength as a function of position (Fig. 1c) show that most of the field energy is confined to the defect region with a mode volume of $V \approx (\lambda_0/n)^3 \approx 0.04 \mu\text{m}^3$, where λ_0 is the resonance wavelength of light in vacuum. This V is a typical value for most of the parameter ranges. Since the intracavity field is a standing wave that oscillates from zero to a maximum every quarter wavelength (see top of Fig. 1c), there is a very limited volume of high field strength in which a SQD must be located if it is to couple strongly.

Photoluminescence (PL) measurements were performed in a temperature-controlled liquid-helium cryostat with internal x – y nanopositioners, essential for stability and the ability to re-find a

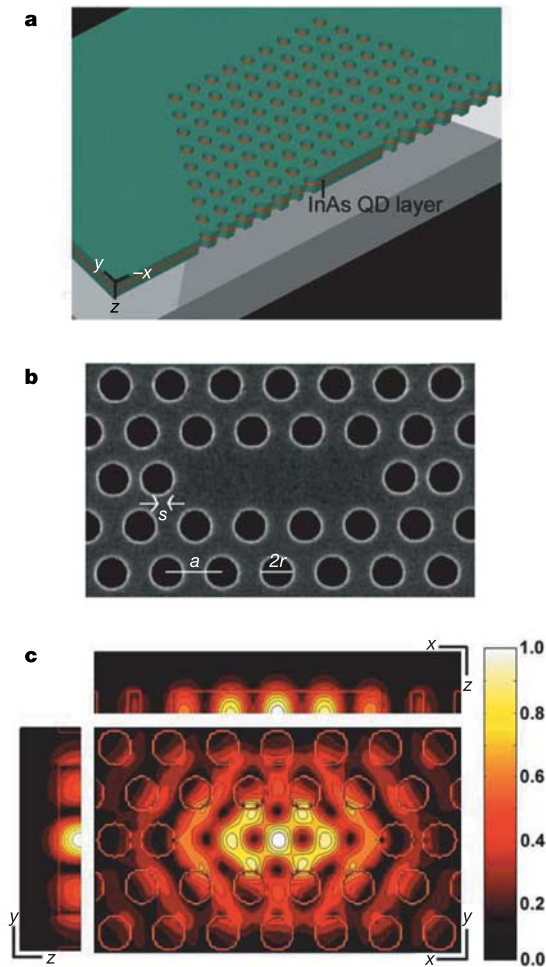


Figure 1 Photonic crystal nanocavity. **a**, Diagram. Half of a hexagon-shaped array of holes forming a nanocavity; scale is $\sim 5 \mu\text{m}$ across at the cut. **b**, Scanning electron micrograph of a fabricated nanocavity, showing the hole spacing a , hole diameter $2r$, and shift s of the two holes at the ends of the 'spacer' formed by omitting three holes. **c**, Computed optical field magnitude superimposed on the nanocavity structure. The scale bar shows the normalized amplitude of the electric field, $|E|/\max(|E|)$. Also shown are a horizontal slice (above main panel) and a vertical slice (left of main panel) through the centre of the spacer.

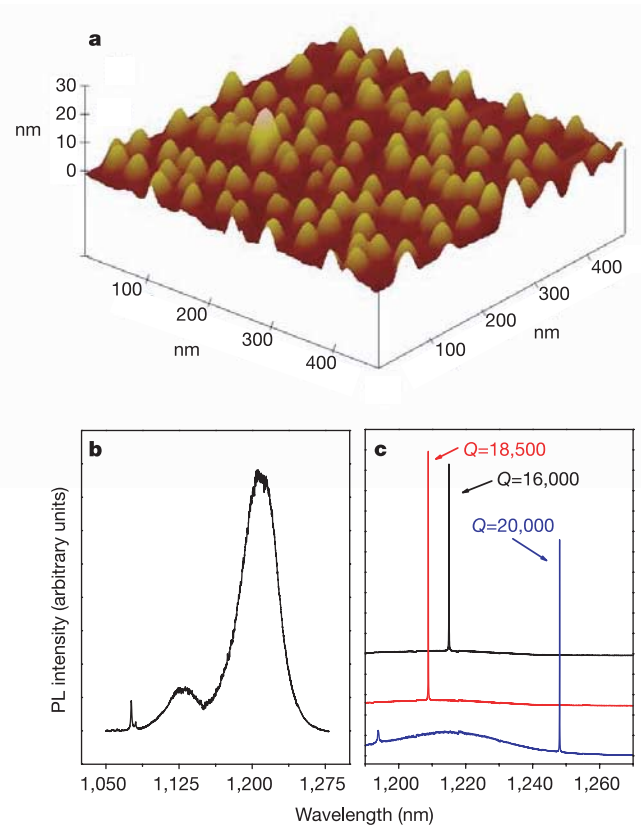


Figure 2 Quantum dots and cavity modes. **a**, Atomic force microscope profile of a layer of InAs QDs similar to the layer used but without layers above. The typical dot size is $\sim 25 \text{ nm}$ diameter and $3\text{--}4 \text{ nm}$ height, and the dot density is $300\text{--}400 \mu\text{m}^{-2}$. **b**, QD ensemble photoluminescence for high excitation power, showing both the lowest ($1,175\text{--}1,250 \text{ nm}$) and first excited ($1,100\text{--}1,150 \text{ nm}$) transitions. **c**, PL from the three nanocavities with the highest values of Q (uncorrected for $\sim 0.04\text{-nm}$ instrument width). Averaging time: 0.1 s for left two peaks; 0.5 s for right.

given nanocavity. The samples were optically pumped by the 770 nm output of a Ti:sapphire continuous wave (c.w.) laser. The pump beam was focused by a reflecting microscope objective (0.5 numerical aperture) to a spot size of $1\text{ }\mu\text{m}$ on the sample. The sample emission was collected by the same microscope objective, analysed with a spectrometer, and detected by an InGaAs array integrating over 0.025 nm per pixel. We estimate that a sample area of $\sim 10\text{ }\mu\text{m}^2$ is imaged into the spectrometer, giving rise to the broad ensemble PL underlying the cavity-related emission in Figs 3 and 4. In this geometry we are using the leakage of the cavity mode out of the top to observe PL from a QD coupled to it. Figure 2b shows the ensemble PL spectrum with the lowest transition line at $\sim 1,200\text{ nm}$, and the first excited transition line at $1,125\text{ nm}$. Figure 2c shows high-power spectra of the three highest- Q nanocavities. There has been a steady improvement in the values of Q obtained for

two-dimensional photonic crystal slab nanocavities^{12,13,26} fabricated for lasers, with a quantum well¹² ($Q = 250$) or ~ 80 quantum dots²⁶ ($Q = 2,000$) as the active medium.

We did not find a SQD coupled strongly to one of the highest- Q modes displayed in Fig. 2c. But a slightly lower- Q ($\sim 13,300$) mode does couple to a SQD located spectrally on the short wavelength side of the lowest energy transition of the ensemble shown in Fig. 2b. At high power (Fig. 3a as in Fig. 2c), the emission is dominated by the cavity peak, because QDs not coupled to the nanocavity are saturated; that is, a QD's emission rate is determined by its radiative decay rate, not by its excitation rate. Therefore, coupled dots emit more photons per unit time than uncoupled dots, owing to Purcell enhancement of spontaneous emission²³. A time-resolved experiment would be needed to see the faster decay of a coupled dot.

At intermediate power ($25\text{ }\mu\text{W}$), the increased QD absorption reduces the Q to 8,000. At low power (Fig. 3b), one can begin to see PL peaks from uncoupled QDs; we note that they all move together in the same way with temperature as does the empty cavity mode, but at a rate much faster than that mode. Therefore, a QD transition can be temperature-scanned through the cavity resonance²⁷. Figure 3b shows an anti-crossing of one QD transition with the cavity mode at $1,182.6\text{ nm}$. The two normal modes repel each other in the vicinity of the crossing of the red and blue lines, showing the temperature dependence of the uncoupled QD and cavity mode resonances, respectively.

In Fig. 3c, the two coupled-system peaks are plotted as a function of temperature on an expanded wavelength scale; for zero detuning where the uncoupled dot and nanocavity resonances are degenerate, the coupled system emission is clearly double-peaked. This anti-crossing behaviour is characteristic of true strong coupling, the regime of reversible exchange of energy back and forth between the SQD and the nanocavity—that is, vacuum Rabi oscillations.

Figure 4a shows an independent scan over a narrower temperature range close to zero detuning. The measured zero detuning vacuum Rabi splitting is $2g = 41\text{ GHz} = 170\text{ }\mu\text{eV} = 0.192\text{ nm}$. Figure 4b displays the zero-detuning emission predicted by an analytic expression²⁸. There is some uncertainty in the values of κ and γ , and even more in the location of the QD relative to the field maximum. For the plot, $g = 20.6\text{ GHz}$; assuming that the dot is in the field maximum, this corresponds to $\mu = 29\text{ D}$ and a radiative lifetime of 1.82 ns (ensemble measurements gave $1\text{--}2\text{ ns}$).

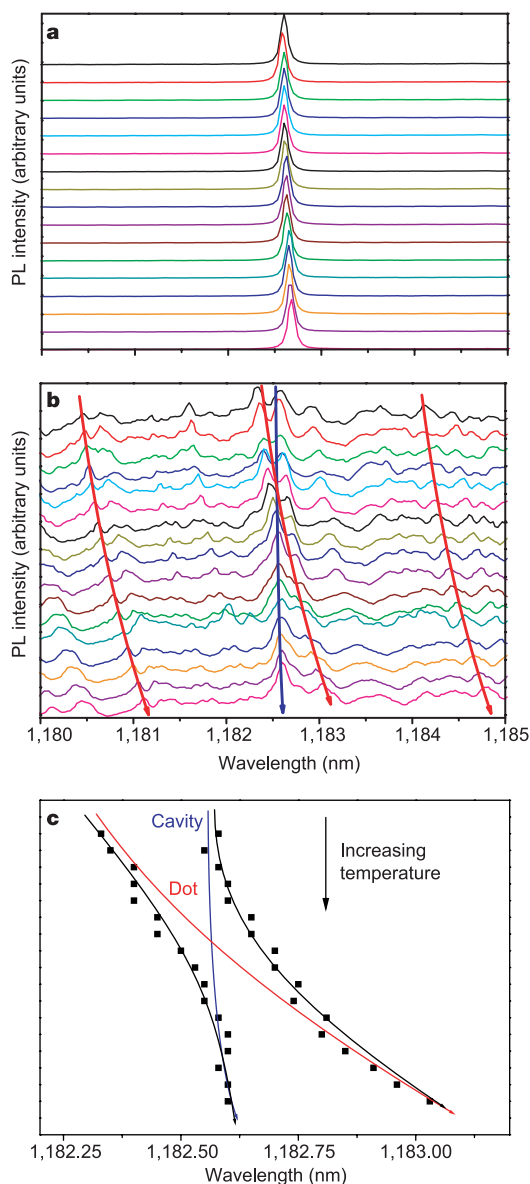


Figure 3 Dot-nanocavity anti-crossing. Temperature is scanned from 13 K at the top to 29 K at the bottom, in 1 K steps. **a**, PL for high excitation power ($690\text{ }\mu\text{W}$) and 0.2 s averaging time; $Q \approx 13,300$. The background QD ensemble emission is $\sim 8\%$ of the peak cavity emission here, and $\sim 50\%$ in **b**. **b**, PL at low power ($0.78\text{ }\mu\text{W}$) and 60 s . **c**, The two coupled-system peaks (black lines are guides for the eye) are plotted as a function of temperature, and compared with the scan rates of an uncoupled QD (red curve) and an empty cavity (blue curve).

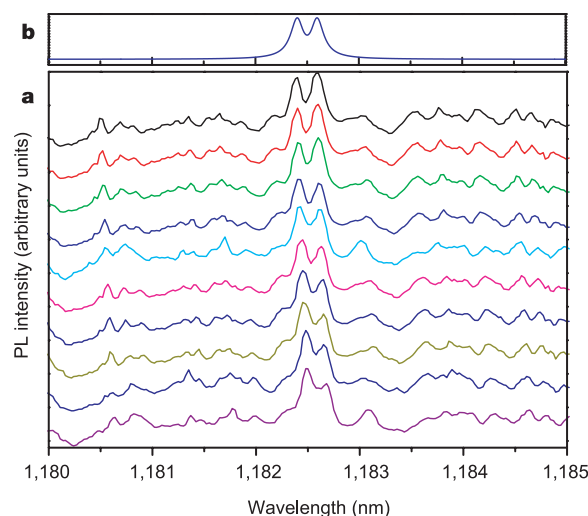


Figure 4 Dot-nanocavity vacuum Rabi splitting. **a**, Near-zero-detuning PL spectra (different run from Fig. 3b), showing double-peaked emission. Temperature is scanned in 0.5 K steps, from 15 K at the top to 19.5 K at the bottom; $0.78\text{ }\mu\text{W}$ and 60 s average. **b**, Plot of analytic expression for zero-detuning emission using $g = 20.6\text{ GHz} = 0.096\text{ nm}$, $\kappa = 42.3\text{ GHz} = 0.197\text{ nm}$, $\gamma = 21.5\text{ GHz} = 0.1\text{ nm}$.

Even though $\sim 10^7$ photons per second are emitted by the coupled system at low power, the signal is weak because most of them stay in the slab; detection through a waveguide coupled to the nanocavity would be far better. If we indeed had only one SQD, there would be little emission from the cavity peak for dot-cavity detunings larger than g ; it occurs because of the high density of weakly coupled QDs. Note that the emission spectrum of a strongly coupled system is double-peaked in all directions, unlike that of a quantum-well planar microcavity, which is double-peaked in the nonperturbative-regime, perpendicular direction while single-peaked in the weak-coupling-regime, in-plane direction²⁹. This means that the energetic position of the emission peaks should be independent of detection direction.

The anti-crossing of Fig. 3b was observed many times by cycling the temperature. In a new sample, we have seen another clear anti-crossing at 1,214.3 nm with vacuum Rabi splitting $2g = 22$ GHz. We have also seen the weak coupling regime of Purcell enhancement of spontaneous emission: temperature scanning causes the QD resonance to cross straight through the cavity resonance, but coupling increases the radiative linewidth.

We expect that our dot/nanocavity system will exhibit truly quantum effects, although the linear spectroscopy that we report here has not proved this. In contrast, even though the normal-mode coupling seen between a single quantum well and a microcavity also results in a two-peaked anti-crossing and is often called strong coupling by the semiconductor community, it is actually semi-classical—much like many-atom vacuum Rabi splitting¹⁰. Even a quantum-well three-dimensional microcavity with a mode diameter of only $\sim 2\text{ }\mu\text{m}$ still requires ~ 300 photons to saturate its vacuum Rabi splitting; thus it is still semi-classical, and far from the quantum regime³⁰.

There are at least two advantages of semiconductor QD cavity QED over atomic cavity QED. First, the dot position is fixed; the ability to do experiments with one and the same quantum emitter is essential for both interesting physics and applications in quantum information science. Second, the ultra-small size of the strongly coupled dot/cavity device, with $>10,000$ cavities mm^{-2} , allows us to speculate about a quantum network that would be able to store, process and distribute quantum information. The interconnections would be made by photonic crystal waveguides (missing lines of holes). The essential element of a quantum network is a deterministic strong coupling of a single dot to a high-finesse optical photonic crystal cavity—as demonstrated here. □

Received 7 August; accepted 19 October 2004; doi:10.1038/nature03119.

- Berman, P. (ed.) *Cavity Quantum Electrodynamics* (Academic, San Diego, 1994).
- Brune, M. *et al.* Quantum Rabi oscillations: A direct test of field quantization in a cavity. *Phys. Rev. Lett.* **76**, 1800–1803 (1996).
- Haroche, S. Entanglement, decoherence, and the quantum/classical boundary. *Phys. Today* 36–42 (July 1998).
- Raimond, J. M., Brune, M. & Haroche, S. *Colloquium: Manipulating quantum entanglement with atoms and photons in a cavity.* *Rev. Mod. Phys.* **73**, 565–582 (2001).
- Mabuchi, H. & Doherty, A. C. Cavity quantum electrodynamics: Coherence in context. *Science* **298**, 1372–1377 (2002).
- McKeever, J., Boca, A., Boozer, A. D., Buck, J. R. & Kimble, H. J. Experimental realization of a one-atom laser in the regime of strong coupling. *Nature* **425**, 268–271 (2003).
- Keller, M., Lange, B., Hayasaka, K., Lange, W. & Walther, H. Deterministic coupling of single ions to an optical cavity. *Appl. Phys. B* **76**, 125–128 (2003).
- Kreuter, A. *et al.* Spontaneous emission lifetime of a single trapped Ca^+ ion in a high finesse cavity. *Phys. Rev. Lett.* **92**, 203002 (2004).
- Wallraff, A. *et al.* Strong coupling of a single photon to a superconducting qubit using circuit quantum electrodynamics. *Nature* **431**, 162–167 (2004).
- Khitrova, G., Gibbs, H. M., Jahnke, E., Kira, M. & Koch, S. W. Nonlinear optics of normal-mode-coupling semiconductor microcavities. *Rev. Mod. Phys.* **71**, 1591–1639 (1999).
- Yablonovitch, E. Inhibited spontaneous emission in solid-state physics and electronics. *Phys. Rev. Lett.* **58**, 2059–2062 (1987).
- Painter, O. *et al.* Two-dimensional photonic band-gap defect mode laser. *Science* **284**, 1819–1821 (1999).
- Akahane, Y., Asano, T., Song, B.-S. & Noda, S. High-Q photonic nanocavity in a two-dimensional photonic crystal. *Nature* **425**, 944–947 (2003).
- Marzin, J.-Y., Gérard, J.-M., Izraël, A., Barrier, D. & Bastard, G. Photoluminescence of single InAs quantum dots obtained by self-organized growth on GaAs. *Phys. Rev. Lett.* **73**, 716–719 (1994).
- Brunner, K., Abstreiter, G., Böhm, G., Tränkle, G. & Weimann, G. Sharp-line photoluminescence and

- two-photon absorption of zero-dimensional biexcitons in a GaAs/AlGaAs structure. *Phys. Rev. Lett.* **73**, 1138–1141 (1994).
- Gammon, D. & Steel, D. G. Optical studies of single quantum dots. *Phys. Today* 36–41 (October 2002).
- Weisbuch, C., Nishioka, M., Ishikawa, A. & Arakawa, Y. Observation of the coupled exciton-photon mode splitting in a semiconductor quantum microcavity. *Phys. Rev. Lett.* **69**, 3314–3317 (1992).
- Gammon, D., Snow, E. S., Shanabrook, B. V., Katzer, D. S. & Park, D. Fine structure in the optical spectra of single GaAs quantum dots. *Phys. Rev. Lett.* **76**, 3005–3008 (1996).
- Petroff, P. M., Lorke, A. & Imamoglu, A. Epitaxially self-assembled quantum dots. *Phys. Today* 46–52 (May 2001).
- Zrenner, A. *et al.* Coherent properties of a two-level system based on a quantum-dot photodiode. *Nature* **418**, 612–614 (2002).
- Moreau, E. *et al.* Single-mode solid-state single photon source based on isolated quantum dots in pillar microcavities. *Appl. Phys. Lett.* **79**, 2865–2867 (2001).
- Bayer, M. *et al.* Inhibition and enhancement of the spontaneous emission of quantum dots in structured microcavities. *Phys. Rev. Lett.* **86**, 3168–3171 (2001).
- Happ, T. D. *et al.* Enhanced light emission of $\text{In}_x\text{Ga}_{1-x}\text{As}$ quantum dots in a two-dimensional photonic-crystal defect microcavity. *Phys. Rev. B* **66**, 041303(R) (2002).
- Vahala, K. J. Optical microcavities. *Nature* **424**, 839–846 (2003).
- Santori, C., Fattal, D., Vučković, J., Solomon, G. S. & Yamamoto, Y. Indistinguishable photons from a single-photon device. *Nature* **419**, 594–597 (2002).
- Yoshie, T., Shchekin, O. B., Chen, H., Deppe, D. G. & Scherer, A. Planar photonic crystal nanolasers (II): Low-threshold quantum dot lasers. *IEICE Trans. Electron.* **E87-C**, 300–307 (2004).
- Kiraz, A. *et al.* Cavity-quantum electrodynamics using a single InAs quantum dot in a microdisk structure. *Appl. Phys. Lett.* **78**, 2932–2934 (2001).
- Carmichael, H. J., Brecha, R. J., Raizen, M. G., Kimble, H. J. & Rice, P. R. Subnatural linewidth averaging for coupled atomic and cavity-mode oscillators. *Phys. Rev. A* **40**, 5516–5519 (1989).
- Stanley, R. P., Houdré, R., Weisbuch, C., Oesterle, U. & Illegems, M. Cavity-polariton photoluminescence in semiconductor microcavities: Experimental evidence. *Phys. Rev. B* **53**, 10995–11007 (1996).
- Lee, E. S. *et al.* Saturation of normal-mode coupling in aluminium-oxide-aperture semiconductor nanocavities. *J. Appl. Phys.* **89**, 807–809 (2001).

Acknowledgements The Caltech group thanks S. Noda and Y. Akahane for discussions on the cavity designs, and the MURI Center for Photonic Quantum Information Systems (ARO/ARDA), NSF-ECS-NIRT and AFOSR for financial support. The Tucson group thanks E. Yablonovich for suggestions, and AFOSR, DURINT, NSF-AMOP and NSF-ECS-EPDT for support. The Texas group acknowledges support from NSF-ECS-NIRT.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to G. K. (galina@optics.arizona.edu).

Current-induced resonance and mass determination of a single magnetic domain wall

Eiji Saitoh¹, Hideki Miyajima¹, Takehiro Yamaoka² & Gen Tatara³

¹Department of Physics, Keio University, Yokohama, 223-8522, Japan

²SII NanoTechnology Inc., Matsudo, 270-2222, Japan

³Department of Earth and Space Science, Osaka University, Osaka, 560-0043, Japan

A magnetic domain wall (DW) is a spatially localized change of magnetization configuration in a magnet. This topological object has been predicted to behave at low energy as a composite particle with finite mass¹. This particle will couple directly with electric currents as well as magnetic fields, and its manipulation using electric currents^{2–8} is of particular interest with regard to the development of high-density magnetic memories⁹. The DW mass sets the ultimate operation speed of these devices, but has yet to be determined experimentally. Here we report the direct observation of the dynamics of a single DW in a ferromagnetic nanowire, which demonstrates that such a topological particle has a very small but finite mass of 6.6×10^{-23} kg. This measurement was realized by preparing a tunable DW potential in the

nanowire, and detecting the resonance motion of the DW induced by an oscillating current. The resonance also allows low-current operation, which is crucial in device applications; a DW displacement of $10\ \mu\text{m}$ was induced by a current density of $10^{10}\ \text{A m}^{-2}$.

The DW mass is, in principle, detectable directly by resonance absorption measurements; magnetic-resonance effects have long been investigated for DWs trapped in pinning centres¹⁰. However, owing to their low sensitivity, these resonance measurements are limited only to samples containing huge numbers of DWs with inevitable distribution. As a result, single DW properties have never been accessible.

To observe the resonance of a single DW, we use DW motion induced by an oscillating electric current in a tunable potential. Recent studies have revealed that a DW can be driven by an electric current by way of several different mechanisms^{2–8}. Theories predict two dominant mechanisms in submicrometre wires¹¹, both arising from the exchange coupling between the local magnetization and the spin of conduction electrons. The first is due to the reflection of conduction electrons by the DW, called momentum transfer; this effect is proportional to the charge current and the DW resistance. The second is the transfer of the spin angular momentum from conduction electrons to the DW as the electrons pass through the DW. This effect, called the spin-transfer effect, is proportional to the spin polarization of the current (spin current). In the case of narrow metallic wires under a steady current, spin transfer is expected to be dominant and recent experiments using d.c. currents^{3–8} are consistent with the above mechanism. The effects of the magnetic field generated by a current (Oersted field) and the hydromagnetic force were shown to be negligibly small in a narrow ferromagnetic wire¹². So far, studies on the current-induced motion have been limited to the effect of a d.c. current or a current pulse. Here we explore a new aspect of the current-driven DW motion using a high-frequency a.c. current. We report two main findings. First, we show that, by tuning the current frequency to the resonance, precise detection of the mass of a single DW is possible. Second, the driving mechanism of the DW motion in the MHz range is clarified, which turns out to be distinct from the d.c. current case.

Figure 1a shows a schematic illustration of the sample system used in the present study. The sample consists of a semicircular loop of soft ferromagnetic $\text{Ni}_{81}\text{Fe}_{19}$ wire with a width of $70\ \text{nm}$, a thickness of $45\ \text{nm}$ and a radius (r) of $50\ \mu\text{m}$, connected to two Cu electrodes (V_+ and V_-). Figure 1b is a magnified view of a scanning electron micrograph of the region surrounding the gap between the electrodes. The complex impedance Z between the electrodes V_+ and V_- is measured by applying an a.c. electric current with a constant amplitude of $I = 100\ \mu\text{A}$ at room temperature. From the Z data, the a.c. resistance $R = \text{Re}Z$ is deduced.

Taking advantage of the semicircular loop structure of the wire, we are able to introduce a single DW in a controllable manner. In the soft ferromagnetic $\text{Ni}_{81}\text{Fe}_{19}$ narrow wire, the magnetization is directed along the wire owing to the strong magnetic-shape anisotropy. A DW is produced by applying a magnetic field $H_{\text{ini}}^y (= 10\ \text{kOe})$ in the direction y to saturate the magnetization, and then decreasing the field to zero (see Fig. 1d). A magnetic-force-microscope (MFM) image in Fig. 2a clearly shows a DW introduced at the bottom of the wire (denoted as α in Fig. 1b) in zero field. The detailed in-plane magnetization configuration around the DW, calculated using a micromagnetic simulator¹³, is shown in Fig. 1c. The remanent state without DWs is prepared using an initialization field of $H_{\text{ini}}^x (= 10\ \text{kOe})$ in the x direction (see Fig. 1e), as seen in a zero-field MFM image (Fig. 2b).

This particular wire structure allows the preparation of a well-defined and tunable DW potential. DWs in the loop have identical exchange energy and magnetic charge $Q_M = 2\mu_0 M_S S$ in zero field, where M_S and S denote the saturation magnetization and the cross-sectional area of the loop, respectively. The measurement is carried

out by applying an external magnetic field H along the direction y . The field is much weaker than the saturation field ($\sim 15\ \text{kOe}$), so the magnetization configuration in the loop is little affected. The motion of a DW confined in the wire loop is then equivalent to that of a 'simple pendulum', where the DW in the wire corresponds to the weight of a pendulum, and the magnetostatic interaction between the magnetic field and the DW magnetic charge corresponds to gravitational force in the pendulum analogy (see Fig. 1f). The DW potential energy consists only of the magnetostatic potential energy (measured from $y = 0$) $U = -Q_M H y \sim -Q_M H(r - (x^2/2r))$ in the vicinity of the bottom of the loop, where x and y represent components of small displacement along the wire. By changing H , we can control the DW-potential curvature and thus the eigenfrequency f_e of the DW according to the following

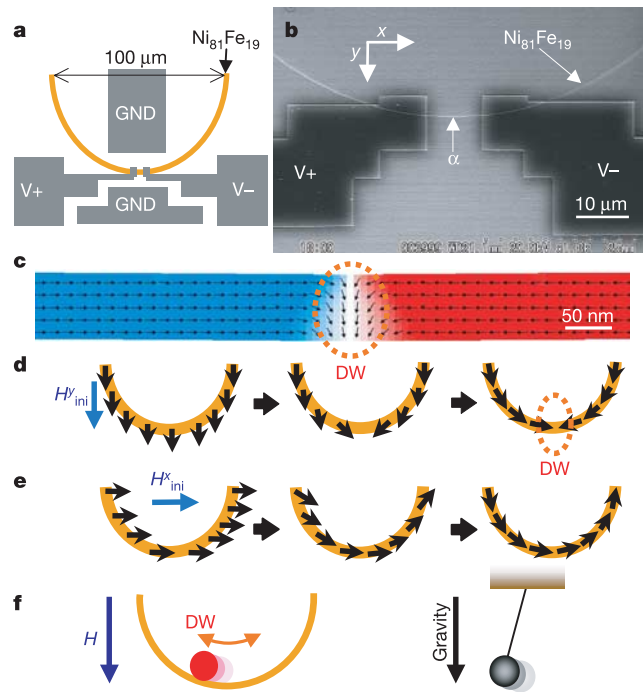


Figure 1 Experimental set-up. **a**, Schematic illustration of the sample system. The system consists of a $\text{Ni}_{81}\text{Fe}_{19}$ semicircular wire loop with a width of $70\ \text{nm}$ and Cu electrodes (V_+ , V_- and GND) fabricated on a thermally oxidized Si substrate. The guard electrodes (GND) are grounded. The whole was prepared by the lift-off technique using electron-beam lithography on a thermally oxidized p-type Si(100) substrate. The sample film is placed between a pair of grounded Cu plates. The impedance Z between electrodes V_+ and V_- is measured by applying an a.c. electric current with a constant amplitude of $I = 100\ \mu\text{A}$, using an a.c. resistance meter (HIOKI3535). The residual impedance component between the resistance meter and the semicircular segment is compensated numerically by measuring the impedance of a similar sample where the $\text{Ni}_{81}\text{Fe}_{19}$ loop is missing and a sample where the electrodes V_+ and V_- are shorted. The capacitance between these electrodes is below $0.2\ \text{pF}$. From the Z data, the a.c. resistance $R = \text{Re}Z$ is deduced. R is simply proportional to the energy dissipation in the wire per unit time in the constant- I mode, while $\text{Im}Z$ is affected also by the self inductance of the wire. In the present measurements, $|\text{Re}Z| \gg |\text{Im}Z|$ for all the current frequencies. **b**, Scanning electron microscope image around the gap in the electrodes V_+ and V_- . The bottom of the $\text{Ni}_{81}\text{Fe}_{19}$ loop is denoted as α . **c**, The in-plane magnetization distribution around an introduced domain wall (DW) at the bottom of the wire loop calculated using a micromagnetic simulator¹³. The arrows represent the local magnetization distribution. A DW is at the centre of the wire. **d, e**, Evolution of magnetization configurations in the $\text{Ni}_{81}\text{Fe}_{19}$ loop after the application of initial magnetic fields H_{ini}^y and H_{ini}^x . The arrows represent the local magnetization distribution. **f**, Schematic illustration of the DW magnetostatic potential energy. This potential energy of a DW in a magnetic field H is equivalent to that of a simple pendulum in a gravity field.

relation:

$$f_e^2 = \frac{Q_M H}{4\pi^2 m r} \quad (1)$$

where m represents the DW mass. When the frequency of the applied a.c. current approaches f_e , the energy dissipation owing to the DW motion is critically enhanced (see below). This enhancement is detected as an increase in the a.c. resistance R , since the energy dissipation in the wire per unit time is equal to RI^2 in the constant- I mode.

Figure 2c shows the frequency(f) spectrum of the a.c. resistance R at $H = 150$ Oe for the $\text{Ni}_{81}\text{Fe}_{19}$ loop segment with a DW present, and Fig. 2d the spectrum in the absence of DWs. We measured these spectra by applying the field H in the direction y . After the introduction of the DW, an unconventional broad peak structure appears at 25 MHz as shown in Fig. 2c, which disappears when the DW is annihilated (Fig. 2d); evidently, the peak structure is of DW origin. In Fig. 3a, the frequency spectra of $\Delta R(H) = R_y(H) - R_x(H)$ for the various values of H below 150 Oe are plotted, where $R_y(H)$ and $R_x(H)$ are the a.c. resistance of the state with and without a DW, prepared by H_{ini}^y and H_{ini}^x , respectively. The peak frequency f_0 of the spectra clearly shifts toward zero as H decreases. Importantly, the squared peak frequency changes almost linearly with H , as shown in Fig. 3b, consistent with equation (1). This is direct evidence that these broad peak structures are due to the resonance absorption by the DW oscillating along the wire around the potential minimum. From the resonance data and equation (1), we estimate the mass of the single DW as $m = (6.55 \pm 0.06) \times 10^{-23}$ kg. This value is comparable to the value $\sim 1 \times 10^{-22}$ kg theoretically given by $m = h^2 N / (4\pi^2 K \lambda^2)$ (refs 1, 11), where h , N , K and λ represent respectively Planck's constant, the number of spins in the DW, the transverse-magnetic-anisotropy energy $\sim 1.8 \times 10^{-24}$ J (estimated

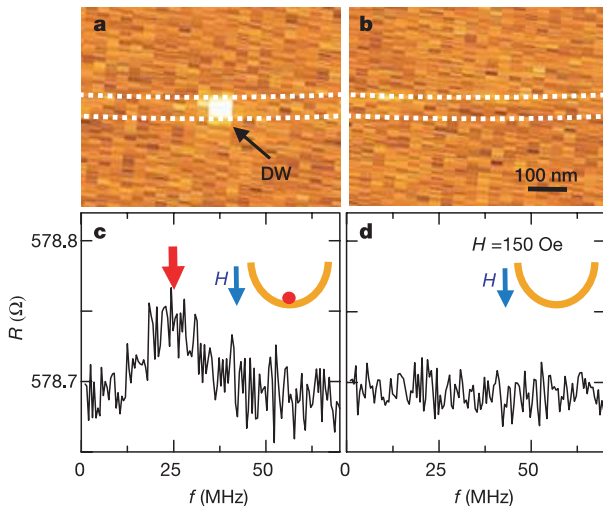


Figure 2 Comparison between experimental results with and without a domain wall (DW). **a, b**, Magnetic-force-microscope images around the bottom of the $\text{Ni}_{81}\text{Fe}_{19}$ loop in remanent magnetic states measured with a scanning probe microscopy system (SPI4000/SPA-300HV, SII NanoTechnology) equipped with a low-moment CoPtCr tip. Before the measurement shown in **a** and **b**, the initial fields $H_{\text{ini}}^y = 10$ kOe and $H_{\text{ini}}^x = 10$ kOe are applied, respectively, which are then set to zero. The dashed lines represent the outlines of the $\text{Ni}_{81}\text{Fe}_{19}$ loop. A DW is imaged as a bright contrast, which corresponds to the stray field from a positive magnetic charge. **c**, Frequency f dependence of the a.c. resistance R for the system with a DW measured by applying an external magnetic field of 150 Oe in the direction y . The arrow represents the frequency at which R reaches a maximum. **d**, Frequency f dependence of the a.c. resistance R for the system without DWs measured by applying an external magnetic field of 150 Oe in the direction y .

by a magnetic torque measurement for an array of $\text{Ni}_{81}\text{Fe}_{19}$ nanowires), and the DW width ~ 70 nm (see Figs 1c and 2a).

The obtained spectral shapes carry the information on the driving mechanism of a DW. The equation of motion¹¹ of a DW for small oscillation is written as $m\ddot{x} = -m(2\pi f_e)^2 x - (m/\tau)\dot{x} + F(t)$, where τ denotes the DW relaxation time. $F(t)$ represents the force due to the current¹¹, which is given in the Fourier representation as $F(f) = enSIR_{\text{DW}}(1 - i\alpha\frac{hf}{K}) - i\frac{h^2}{2\pi eK\lambda}fI_s$, where n and R_{DW} represent conduction-electron density and the DW resistance, respectively. The spin current, I_s , is proportional to I . The energy dissipation rate due to the a.c.-current-driven DW motion is calculated from this equation of motion. This rate should be equal to the difference of the energy dissipation per unit time in the states with and without DW, $\Delta R I^2$. We thus obtain:

$$\Delta R(f) = \frac{1}{8\pi^2 m \tau I^2} \frac{f^2 |F(f)|^2}{(f^2 - f_e^2)^2 + (f/(2\pi\tau))^2} \quad (2)$$

We found that this expression fits quite well to the observed spectra using only the momentum-transfer term if we choose $\tau = (1.4 \pm 0.2) \times 10^{-8}$ s and $R_{\text{DW}} = (2.6 \pm 0.7) \times 10^{-4}$ Ω for all the H values as exemplified in Fig. 3c. (Note that τ and R_{DW} are the only fitting parameters here.) What is notable here is that the spectra cannot be reproduced if we assume spin transfer dominates, that is, if $F \propto f$, in which case the resonance-peak structure should become

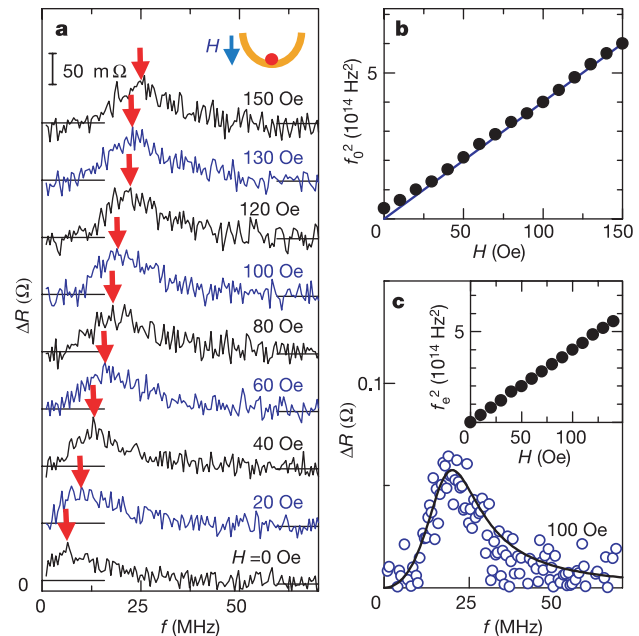


Figure 3 External-field dependence of the a.c. resistance spectra. **a**, Frequency f spectra of $\Delta R(H) = R_y(H) - R_x(H)$ for various external fields H , where $R_y(H)$ and $R_x(H)$ denote the a.c. resistance with and without a domain wall (DW), prepared by the initial fields H_{ini}^y and H_{ini}^x , respectively. The applied magnetic field during the measurement is directed along y . The arrows represent frequencies at which ΔR reaches maxima in each spectrum. **b**, External field H dependence of the squared peak frequencies (f_0^2) of ΔR spectra shown by arrows in **a**. **c**, An example of the fitting result (a solid curve) for the measured ΔR spectrum at $H = 100$ Oe (open circles) using ΔR calculated in the small-oscillation limit. The squared resonance frequencies (f_0^2) determined by the fitting show excellent linear dependence on H (see inset), and the DW mass estimated using these data is identical to that estimated from the data shown in **b**. We note that the analysis here does not reduce to the d.c. case in the limit of $f \rightarrow 0$, since it is assumed here that the deviation of the magnetization from the easy plane is small, which linearizes the equation of motion for x . This assumption does not hold in most d.c. cases, where large displacement of a DW is discussed and the deviation of the magnetization from the easy plane can be large¹¹.

asymmetric. Moreover, if using only the spin-transfer term, we need to assume an unreasonable spin-current density much larger than the current density. The above values of τ and R_{DW} indicate that the contribution of spin transfer to F is two orders of magnitude smaller than that of momentum transfer. This is quite general when the small oscillation of a DW is slow compared with the microscopic dynamics of the single spin¹⁴, since the spin transfer force has a suppression factor of $\hbar/K \approx 0.01$. A striking point here is that although R_{DW} is very small, the momentum transfer force is amplified significantly at resonance, and becomes detectable.

The present DW motion is driven by an extremely small current density, $3 \times 10^{10} \text{ A m}^{-2}$, which is two orders of magnitude smaller than the d.c. cases⁷. Because of this small current density, the heating effect is negligibly small in the measurement. Even for this small current, the displacement of the DW is estimated to be $\Delta x \approx 10 \mu\text{m}$. Thus the current required for $\Delta x \approx 1 \mu\text{m}$ is $3 \times 10^9 \text{ A m}^{-2}$. This suggests the possibility of efficient writing in magnetic memory devices using a.c.-current-induced depinning and d.c.-current-driven drift of a DW. Very recently, Lim *et al.*¹⁵ reported that the use of fast current pulses of duration down to 400 ps allowed them to drive a DW at very low current densities. This result may be understood in terms of the resonance driving reported here, since such fast pulses comprise not only a d.c. component but also many high-frequency components.

The resonance-amplified spectroscopy presented here would be a powerful probe of nanoscale magnetism in the quantum regime—for example, quantum tunnelling of DWs. The resonance observed here may also be useful as a way to produce magnetoresistance under an a.c. current at a fixed frequency. An external field shifts the resonance frequency as shown in Fig. 3a: the field drives a system that is in-resonance to be out of resonance (and vice versa). Thus, large a.c. magnetoresistance both with positive and negative signs can be realized by preparing a system with large τ and R_{DW} . A proper choice of materials and reducing the DW width by narrowing the wire will increase τ and R_{DW} , respectively. □

Received 25 May; accepted 10 September 2004; doi:10.1038/nature03009.

1. Döring, V. W. Über die Trägheit der Wände zwischen Weisschen Bezirken. *Z. Naturforsch.* **3a**, 373–379 (1948).
2. Berger, L. Exchange interaction between ferromagnetic domain wall and electric current in very thin metallic films. *J. Appl. Phys.* **55**, 1954–1956 (1984).
3. Freitas, P. P. & Berger, L. Observation of s-d exchange force between domain walls and electric current in very thin Permalloy films. *J. Appl. Phys.* **57**, 1266–1269 (1988).
4. Hung, C.-Y. & Berger, L. Exchange forces between domain wall and electric current in permalloy films of variable thickness. *J. Appl. Phys.* **63**, 4276–4278 (1988).
5. Gan, L. *et al.* Pulsed-current-induced domain wall propagation in Permalloy patterns observed using magnetic force microscope. *IEEE Trans. Magn.* **36**, 3047–3049 (2000).
6. Koo, H., Krafft, C. & Gomez, R. D. Current-controlled bi-stable domain configurations in $\text{Ni}_{81}\text{Fe}_{19}$ elements: An approach to magnetic memory devices. *Appl. Phys. Lett.* **81**, 862–864 (2002).
7. Yamaguchi, A. *et al.* Real-space observation of current-driven domain wall motion in submicron magnetic wires. *Phys. Rev. Lett.* **92**, 077205 (2004).
8. Grollier, J. *et al.* Switching a spin valve back and forth by current-induced domain wall motion. *Appl. Phys. Lett.* **83**, 509–511 (2003).
9. Yamanouchi, M., Chiba, D., Matsukura, F. & Ohno, H. Current-induced domain-wall switching in a ferromagnetic semiconductor structure. *Nature* **428**, 539–542 (2004).
10. Rado, G. T., Wright, R. W., Emerson, W. H. & Terris, A. Ferromagnetism at very high frequencies IV: Temperature dependence of the magnetic spectrum of a ferrite. *Phys. Rev.* **88**, 909–915 (1952).
11. Tataru, G. & Kohno, H. Theory of current-driven domain wall motion: Spin transfer versus momentum transfer. *Phys. Rev. Lett.* **92**, 086601 (2004).
12. Berger, L. Low-field magnetoresistance and domain drag in ferromagnets. *J. Appl. Phys.* **49**, 2156–2161 (1978).
13. Donahue, M. J. & Porter, D. G. The Object Oriented MicroMagnetic Framework (OOMMF) project at ITL/NIST. (<http://math.nist.gov/oommf/>) (1998).
14. Chikazumi, S. *Physics of Ferromagnetism* 2nd edn 559–561 (Oxford Univ. Press, Oxford, 1997).
15. Lim, C. K. *et al.* Domain wall displacement induced by subnanosecond pulsed current. *Appl. Phys. Lett.* **84**, 2820–2822 (2004).

Acknowledgements This work was supported by a Grant-in-Aid from the Ministry of Education, Culture, Sports, Science, and Technology, Japan. G.T. thanks The Mitsubishi Foundation for financial support.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to E.S. (eizi@phys.keio.ac.jp).

A fast low-power optical memory based on coupled micro-ring lasers

Martin T. Hill¹, Harmen J. S. Dorren¹, Tjibbe de Vries¹, Xaveer J. M. Leijtens¹, Jan Hendrik den Besten¹, Barry Smalbrugge¹, Yok-Siang Oei¹, Hans Binsma², Giok-Djan Khoe¹ & Meint K. Smit¹

¹COBRA Research Institute, Technische Universiteit Eindhoven, Postbus 513, 5600 MB Eindhoven, The Netherlands

²JDS Uniphase, Prof. Holstlaan 4, 5656 AA Eindhoven, The Netherlands

The increasing speed of fibre-optic-based telecommunications has focused attention on high-speed optical processing of digital information¹. Complex optical processing requires a high-density, high-speed, low-power optical memory that can be integrated with planar semiconductor technology for buffering of decisions and telecommunication data². Recently, ring lasers with extremely small size and low operating power have been made^{3–7}, and we demonstrate here a memory element constructed by interconnecting these microscopic lasers. Our device occupies an area of $18 \times 40 \mu\text{m}^2$ on an InP/InGaAsP photonic integrated circuit, and switches within 20 ps with 5.5 fJ optical switching energy. Simulations show that the element has the potential for much smaller dimensions and switching times. Large numbers of such memory elements can be densely integrated and interconnected on a photonic integrated circuit: fast digital optical information processing systems employing large-scale integration should now be viable.

Micro-resonators are widely seen as a promising approach for low-power nonlinear integrated photonic devices. An optical memory or bistable element has been demonstrated⁸ which employed a passive planar micro-resonator of similar size to those presented here. High optical operating power induced a thermal bistability, which had microsecond switching times. Logic operations using the same resonators with optical pulses to avoid thermal effects showed picojoule switching energies⁸. Simulations⁹ indicate that the semiconductor-based passive planar micro-resonator approach may eventually achieve switching energies of hundreds of femtojoules, with micrometre dimensions and switching times of the order of 10 ps. For comparison with electronics, switching energies of 1 fJ or less occur in recent silicon large-scale integration (LSI) circuits¹⁰, albeit with slower switching speed. Memory elements employing large bistable lasers have achieved low switching energies¹¹. But until now, no concept has been demonstrated that permits micrometre-sized planar lasers to function as memory elements, as these lasers tolerate only limited coupling to the external environment. Here we employ the inherent properties of ring lasers to achieve the strong interactions between lasers that are necessary for bistable operation. The resulting optical memory element can be made with micrometre dimensions, unlimited memory lifetime and picosecond switching times, while having switching energies several orders of magnitude less than those of passive micro-resonator systems^{8,9}. Furthermore, the static operating power required per memory element can be in the microwatt range^{3–7}, three orders of magnitude less than passive micro-resonators⁹.

Micro-ring lasers typically have two inherent lasing modes; laser light travelling in the clockwise (CW) direction, and laser light in the anticlockwise (ACW) direction. Furthermore, in an ideal ring laser, light in the CW mode is not coupled into the ACW mode, and vice versa. These two ring laser properties are exploited to create a system with two stable states, by connecting two ring lasers together via a waveguide (Fig. 1). In state A, CW light from laser A is injected via the waveguide into laser B. The light from laser A will undergo significant resonant amplification in laser B if the

resonant frequencies of the two laser cavities are close. This injected light competes with the laser B self-oscillations for available power from the laser gain medium. If sufficient light is injected into laser B, then the laser B gain will be decreased below threshold. This extinguishes the laser B self-oscillation, and laser A captures or injection-locks¹² laser B, forcing light to circulate only in the CW direction. Only small amounts of light need enter the laser cavity to achieve injection locking, owing to the strong resonant amplification. The reverse situation with laser B injection-locking laser A is also a stable state of the system (Fig. 1b) when the lasers are equally pumped.

To set the system in one state or another, light close to the lasing wavelength and polarization can be injected into the waveguide connecting the lasers. This light will set both lasers simultaneously lasing in either the CW or ACW direction. The different states can be distinguished by the different power levels at the two outputs. The power level at the output associated with the locked laser will be three times that of the other output. This difference occurs because all the pump energy in the locked laser is forced into one lasing mode. Testing if one of the lasers is unidirectionally injection-locked provides a higher contrast output, and can be achieved by coupling an additional waveguide to one of the lasers. Additionally, the lasing wavelengths of the lasers may be different, allowing the states to be distinguished by the wavelength of the light output.

The minimum power coupling κ between a laser mode and the inter-laser waveguide that is required for correct memory operation is now determined. A concise derivation of κ is possible if the laser has a homogeneously broadened gain medium with zero linewidth enhancement factor¹³, and a flat gain spectrum over the wavelengths of interest. Furthermore, lasing occurs equally in CW and ACW modes and there is no coupling between the modes, in an isolated ring laser. These ideal characteristics can be approached in actual ring lasers with well designed cavities and various gain mediums. Departures from the ideal characteristics are tolerable, although

they complicate the model. In particular, a small coupling between the modes, due to ring imperfections, is tolerable, providing it is much smaller than the inter-laser coupling.

The ratio of circulating optical power P_1 in the ring laser due to amplification of light from another laser P_0 is as follows when lasing is still present in the ring¹².

$$\frac{P_1}{P_0} = \frac{\kappa}{4\sin^2 \frac{\phi}{2}} \quad (1)$$

Here ϕ represents the detuning between the two laser cavities; it is given by $\phi = 2\pi n L_B / \lambda_A \bmod 2\pi$, where L_B is the laser B cavity round-trip length, λ_A is the wavelength of the laser A cavity resonance peak, and n the effective refractive index in the cavity. The circulating optical power in each mode is denoted P_L , with a power (in W) of κP_L being coupled out of the ring laser in each direction. For one laser to completely suppress lasing in the other laser requires $P_1 > 2P_L$ when $P_0 = \kappa P_L$. Substituting these values of P_1 and P_0 into equation (1) gives the condition for minimum κ :

$$\kappa > 2\sqrt{2} \left| \sin \frac{\phi}{2} \right| \quad (2)$$

From equation (2) it can be seen that for laser resonators with closely matched resonant frequencies, the required coupling κ can approach zero. This fact permits the use of micro-ring lasers that can tolerate only a moderate out-coupling of light.

To demonstrate the memory concept, micro-ring lasers with an outside diameter of 16 μm were fabricated using InP/InGaAsP material (Fig. 2). The ring and waveguide structures were defined with contact lithography, and were deeply reactive ion etched 800 nm below the light guiding layer to permit small-diameter ring lasers. The ring and inter-laser ridge waveguide width was 2 μm , owing to process limitations. Optical mode phase matching between the ring and inter-laser waveguide was achieved by employing a semicircular inter-laser waveguide¹⁴. The required gap of several hundred nanometres between waveguides was smaller than our lithographic resolution. Therefore we chose to partially fill the gap, which ensured sufficient coupling but introduced high coupling losses. Low loss and reproducible coupling between resonators and waveguides have been predicted and demonstrated using various techniques^{8,14–16}, which are compatible with the resolution of state-of-the-art commercial projection lithography. The ring lasers were fabricated in small active regions on the InP/InGaAsP wafer, which contained an active layer of bulk InGaAsP material with an emission wavelength of 1.55 μm . The

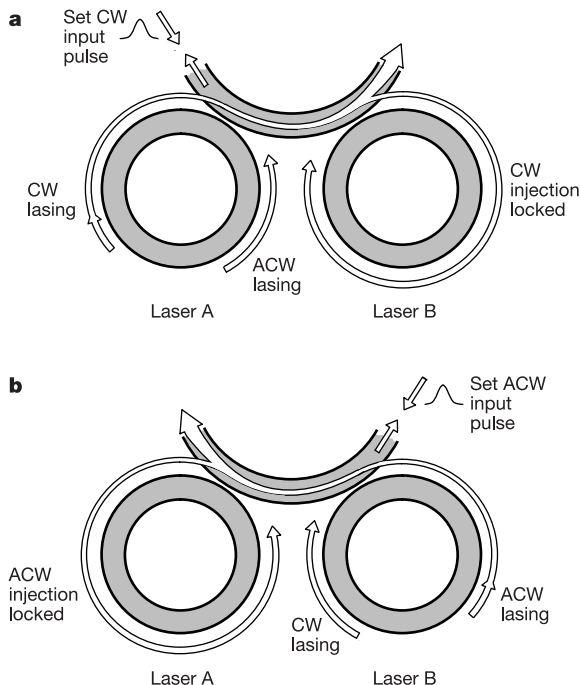


Figure 1 Two micro-ring lasers coupled via a waveguide. The clockwise (CW) and anticlockwise (ACW) lasing modes and optical interaction between the lasers are shown. The system can have two stable states under certain conditions. **a**, In one state, light from laser A injection-locks laser B, forcing it to lase only in the CW direction. **b**, In the second state, laser B injection-locks laser A, forcing it to lase only in the ACW direction. Pulses of light at the chosen input can set the system in the corresponding state.

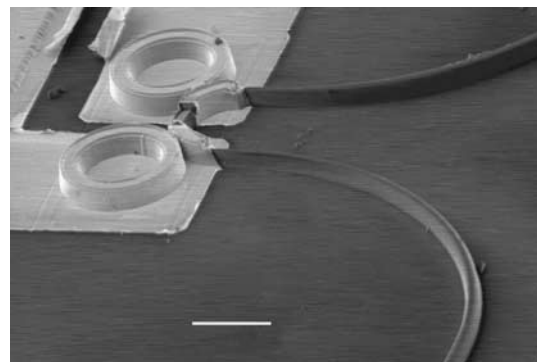


Figure 2 A memory element formed by two 16 μm diameter micro-ring lasers coupled via a waveguide on a InP/InGaAsP photonic integrated circuit. Scale bar, 10 μm . The micro-ring lasers were fabricated in active areas of the integrated circuit containing bulk 1.55 μm bandgap InGaAsP in the light guiding layer. Separate electrical contacts allowed each laser's wavelength to be individually tuned by adjusting the laser current. Passive waveguides connected the micro-ring lasers to the integrated circuit edges.

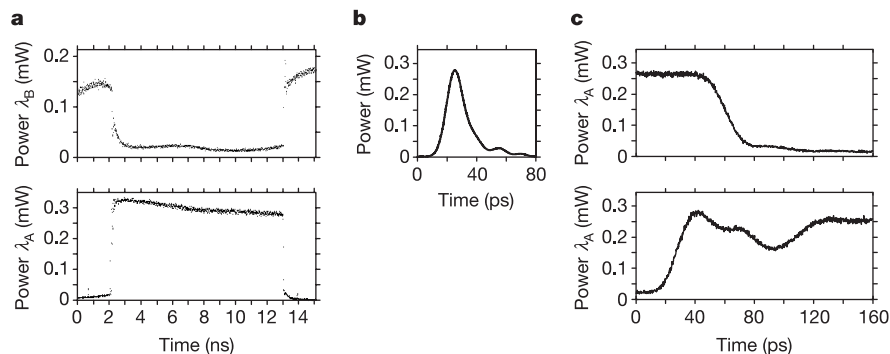


Figure 3 Oscilloscope traces showing the switching of the memory element between states. Light pulses of width (full-width at half-maximum) 13 ps and 5.5 fJ average energy switched the memory state. **a**, Initially the memory is in the state of laser B injection-locking laser A, and the memory outputs light at the laser B wavelength, $\lambda_B = 1,453.5$ nm. At time 2 ns, a light pulse at the laser A wavelength

$\lambda_A = 1,464.6$ nm sets the state to laser A injection-locking laser B, and light at wavelength λ_A is now output. Another light pulse of wavelength λ_B 11 ns later resets the memory. **b**, Oscilloscope trace of the switching pulse for the λ_A low to high transition. **c**, Detail of the on and off transitions for the λ_A wavelength.

ring lasers were connected to the chip edges by waveguides which passed through passive regions, where the light guiding layer consisted only of InGaAsP with a bandgap of 1.25 μm . Details of the wafer growth procedure, layer stack, and waveguide and device fabrication can be found elsewhere¹⁷.

The lasers were operated at 281 K in pulsed mode, with current pulses of 80 ns duration every 13.5 μs . The threshold current for these conditions was approximately 30 mA. The high threshold current and pulsed mode operation were due to losses in the ring to inter-laser waveguide coupling. The coupling junction caused a 50% loss of light in the ring, while only transferring approximately 1.5% of the light in the ring to the inter-laser waveguide, for reasons described above. A d.c. current bias was also applied to one laser to tune its resonant frequency very close to that of the other laser. In operation, the lasers were biased at twice the threshold current. During the current pulse laser B lased at a wavelength of 1,453.5 nm, λ_B . Laser A lased at 1,464.6 nm, λ_A , which was one resonator free spectral range away from λ_B .

Initially the system was set in the laser B dominant state (Fig. 3a). Light emanating from the laser A side of the device was filtered and recorded via a photodiode and oscilloscope, after optical amplification. A 13 ps full-width at half-maximum (FWHM) light pulse (Fig. 3b), with centre frequency 1,464.7 nm, was sent into the laser A side of the device. This pulse caused the system to change state to laser A dominant. 11 ns later, a 13 ps FWHM light pulse with centre frequency 1,453.7 nm was injected into the laser B side of the device. As can be seen from Fig. 3a, the system reverted to the laser B dominant state. Details of the state transitions are shown for the λ_A wavelength in Fig. 3c. The vertical axes of the graphs show power at the input/output of the memory element. The uncertainty in the memory element output power level is approximately $\pm 15\%$, due to filter and optical amplifier polarization dependences. Reference waveguide structures were employed to determine the optical coupling losses from the test set-up to the memory element. It can be seen that the 10 to 90% transition times are less than 20 ps. The high to low and low to high transitions required switching pulse energies of 6.9 fJ and 4 fJ respectively, giving an average 5.5 fJ required for a state transition. The state could be reliably switched with lower energy, 2 fJ, but with approximately 25% longer transition times. The on/off contrast ratio away from the transition regions for this initial realization was at least 8 dB for λ_B and 10 dB for λ_A , a usable level in digital signal processing systems¹⁸.

We believe that considerably improved performance is possible given a more advanced laser gain medium, smaller laser cavities, and efficient laser to waveguide coupling. Quantum confined semiconductor materials are better laser gain mediums owing to their lower

threshold currents^{19,20} and linewidth enhancement factor. Furthermore, they possess nonlinear gain effects^{19,20} caused by finite carrier capture times, which can rapidly exchange the CW and ACW photon populations in the lasers and suppress laser relaxation oscillations²⁰. In the experiment, some effects of the relaxation oscillation can be seen in the turn-on transition (Fig. 3c), where there is a dip in the output after the initial turn on. Smaller laser cavities require less gain medium and so less device pumping power. They also permit a faster build up of switching light inside the laser. Simulations performed for a memory element employing 3 μm diameter disk lasers and quantum well material indicate that switching times of the order of 1 ps with a switching energy less than 1 fJ are possible. Photonic bandgap defect cavity lasers can have an optical mode size less than 0.5 μm in diameter²¹, and also exhibit whispering gallery mode behaviour²², as in disk lasers. The small size of these lasers may allow even smaller switching times and powers. $\square$

Received 10 May; accepted 20 September 2004; doi:10.1038/nature03045.

1. Cotter, D. *et al.* Nonlinear optics for high speed digital information processing. *Science* **286**, 1523–1528 (1999).
2. Zimmerman, S., Wixforth, A., Kotthaus, J. P., Wegscheider, W. & Bichler, M. A semiconductor-based photonic memory cell. *Science* **283**, 1292–1295 (1999).
3. Fujita, M., Ushigome, R. & Baba, T. Continuous wave lasing in GaInAsP microdisk injection laser with threshold current of 40 μA . *Electron. Lett.* **36**, 790–791 (2000).
4. Spillane, S. M., Kippenberg, T. J. & Vahala, K. J. Ultralow-threshold Raman laser using a spherical dielectric microcavity. *Nature* **415**, 621–623 (2002).
5. Polman, A., Min, B., Kalkman, J., Kippenberg, T. J. & Vahala, K. J. Ultralow-threshold erbium-implanted toroidal microlaser on silicon. *Appl. Phys. Lett.* **84**, 1037–1039 (2004).
6. Levi, A. F. J., McCall, S. L., Pearton, S. J. & Logan, R. A. Room temperature operation of submicrometre radius disk laser. *Electron. Lett.* **29**, 1666–1667 (1993).
7. Baba, T. & Sano, D. Low-threshold lasing and Purcell effect in microdisk lasers at room temperature. *IEEE J. Select. Top. Quant. Electron.* **9**, 1340–1346 (2003).
8. Van, V. *et al.* Optical signal processing using nonlinear semiconductor microring resonators. *IEEE J. Select. Top. Quant. Electron.* **8**, 705–713 (2002).
9. Yanik, M. F., Fan, S. & Soljacic, M. High-contrast all-optical bistable switching in photonic crystal microcavities. *Appl. Phys. Lett.* **83**, 2739–2741 (2003).
10. Keyes, R. W. Fundamental limits of silicon technology. *Proc. IEEE* **89**, 227–239 (2001).
11. Kawaguchi, H. *Bistabilities and Nonlinearities in Laser Diodes* (Artech House, Boston, 1994).
12. Buczek, C. J., Freiberg, R. J. & Skolnick, M. L. CO₂ regenerative ring power amplifiers. *J. Appl. Phys.* **42**, 3133–3137 (1971).
13. Harder, C., Vahala, K. & Yariv, A. Measurement of the linewidth enhancement factor α of semiconductor lasers. *Appl. Phys. Lett.* **40**, 328–330 (1983).
14. Sakai, A. & Baba, T. FDTD simulation of photonic devices and circuits based on fan-shaped microdisks. *J. Lightwave Technol.* **17**, 1493–1499 (1999).
15. Hagness, S. C., Rafizadeh, D., Ho, S. T. & Taflov, A. FDTD microcavity simulations: design and experimental realization of waveguide-coupled single-mode ring and whispering-gallery-mode disk resonators. *J. Lightwave Technol.* **15**, 2154–2164 (1997).
16. Noda, S., Chutinan, A. & Imada, M. Trapping and emission of photons by a single defect in a photonic bandgap structure. *Nature* **407**, 608–610 (2000).
17. den Besten, J. H. *et al.* A compact digitally tunable seven-channel ring laser. *IEEE Photon. Technol. Lett.* **14**, 753–755 (2002).
18. Smith, S. D., Walker, A. C., Tooley, F. A. P. & Wherrett, B. S. The demonstration of restoring digital optical logic. *Nature* **325**, 27–31 (1987).

19. Ledentsov, N. N. *et al.* Quantum-dot heterostructure lasers. *IEEE J. Select. Top. Quant. Electron.* **6**, 439–451 (2000).
20. Chuang, S. L. *Physics of Optoelectronic Devices* (Wiley-Interscience, New York, 1995).
21. Painter, O. *et al.* Two-dimensional photonic band-gap defect mode laser. *Science* **284**, 1819–1821 (1999).
22. Ryu, H.-Y. *et al.* Square-lattice photonic band-gap single-cell laser operating in the lowest-order whispering gallery mode. *Appl. Phys. Lett.* **80**, 3883–3885 (2002).

Acknowledgements This work was supported by the Netherlands Organization for Scientific Research (NWO) through the 'NRC photonics' grant.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to M.H. (m.t.hill@tue.nl).

Prolonged KREEP magmatism on the Moon indicated by the youngest dated lunar igneous rock

Lars E. Borg¹, Charles K. Shearer¹, Yemane Asmerom² & James J. Papike¹

¹Institute of Meteoritics, Department of Earth and Planetary Sciences, and

²Department of Earth and Planetary Sciences, University of New Mexico, Albuquerque, New Mexico 87131, USA

Primordial solidification of the Moon (or its uppermost layer) resulted in the formation of a variety of rock types that subsequently melted and mixed to produce the compositional diversity observed in the lunar sample suite^{1,2}. The initial rocks to crystallize from this Moon-wide molten layer (the magma ocean) contained olivine and pyroxene and were compositionally less evolved than the plagioclase-rich rocks that followed. The last stage of crystallization, representing the last few per cent of the magma ocean, produced materials that are strongly enriched in incompatible elements including potassium (K), the rare earth elements (REE) and phosphorus (P)—termed KREEP^{3–5}. The decay of radioactive elements in KREEP, such as uranium and thorium, is generally thought to provide the thermal energy necessary for more recent lunar magmatism^{4,6,7}. The ages of KREEP-rich samples are, however, confined to the earliest periods of lunar magmatism between 3.8 and 4.6 billion years (Gyr) ago^{8,9}, providing no physical evidence that KREEP is directly involved in more recent lunar magmatism. But here we present evidence that KREEP magmatism extended for an additional 1 Gyr, based on analyses of the youngest dated lunar sample.

Northwest Africa 773 (NWA 773) is a 633-g lunar meteorite composed of an impact breccia. The bulk of the sample is an olivine cumulate clast that is interpreted to be of igneous origin^{10,11}. The clast contains approximately 48% olivine, 27% pigeonite, 11% augite, 2% hypersthene, and 11% plagioclase, as well as a mesostasis component characterized by trace amounts of barium K-feldspar, merrillite, troilite, Cr-spinel, and Fe–Ni metal. In addition to these igneous phases, NWA 773 contains secondary alteration products that were added to the meteorite during terrestrial weathering in the African desert. The composition of both the bulk sample and parental melts calculated from mineral compositions demonstrate strong enrichments of incompatible elements relative to mare basalts. This geochemical signature closely matches that of KREEP, and has led to the conclusion that NWA 773 is a mixture of a magnesian magma and an extremely evolved KREEP-rich component¹¹. Thus, NWA 773 has geochemical affinities to other

KREEP-rich samples such as the Mg-suite, alkali suite and the KREEP basalts.

We have completed Rb–Sr and Sm–Nd isotopic analyses on whole-rock and mineral fractions from the olivine clast in NWA 773. We obtained 100 mg of NWA 773 from the Natural History Museum, London. Our analytical procedures are described in the Methods section and are similar to those described by us previously¹². The isotopic results are presented in Table 1 and Fig. 1. The whole-rock and mineral fractions (olivine, olivine + pyroxene and plagioclase) define a Sm–Nd isochron with an age of 2.865 ± 0.031 Gyr, using the Isoplot program¹³. Regression of the data resulted in a mean-squared weighted derivative (MSWD) of about one, indicating that all points lie within uncertainty of the isochron. As a result, the Sm–Nd age appears to be robust.

Our Sm–Nd age is concordant with the Ar–Ar age of ~ 2.91 Gyr reported for NWA 773 (ref. 14). However, whereas Ar–Ar ages commonly reflect resetting events associated with the formation of large lunar impact basins, Sm–Nd ages are likely to represent crystallization ages. Thus, the age of NWA 773 is confirmed to be the youngest crystallization age derived from any lunar sample, as previously included from the Ar–Ar age (ref. 14). Although ages of mare basalt units have been estimated to be as young as 1.2 Gyr from crater size–frequency distribution measurements¹⁵, the youngest crystallization age previously determined for a lunar sample is ~ 3.1 Gyr (ref. 16), whereas the youngest crystallization of a KREEP-rich sample (labelled 15382) reported previously was 3.83 ± 0.02 Gyr ago¹⁷. The 2.865 ± 0.031 -Gyr crystallization age of NWA 773 reported here therefore extends the range of dated lunar samples by about 250 million years (Myr), and the period of KREEP-magmatism by about 1 Gyr.

The Sm–Nd isochron defines an initial ϵ_{Nd} isotopic composition of -7.84 ± 0.22 ($\epsilon_{\text{Nd}} = [^{143}\text{Nd}/^{144}\text{Nd}_{\text{sample}} \div ^{143}\text{Nd}/^{144}\text{Nd}_{\text{CHUR at 2.865 Gyr ago}} - 1] \times 10^4$), indicating the sample is derived from a strongly light-REE (LREE)-enriched source region (CHUR, chondritic uniform reservoir). The $^{147}\text{Sm}/^{144}\text{Nd}$ ratio calculated for the NWA 773 source region is 0.158, and is significantly more LREE-enriched than the source region for any other dated lunar rock (Fig. 2). Shih *et al.*¹⁸ suggested that many samples with KREEP-rich geochemical signatures have Sm–Nd isotopic systematics indicative of derivation from a common source region ($^{147}\text{Sm}/^{144}\text{Nd}$ ratio is 0.181; Fig. 2). Northwest Africa 773 does not lie on this array and is therefore derived from the most evolved source region yet known (that is, with the largest KREEP component). Interestingly, the $^{147}\text{Sm}/^{144}\text{Nd}$ ratio of the NWA 773 source is very similar to the ratio estimated for the KREEP source

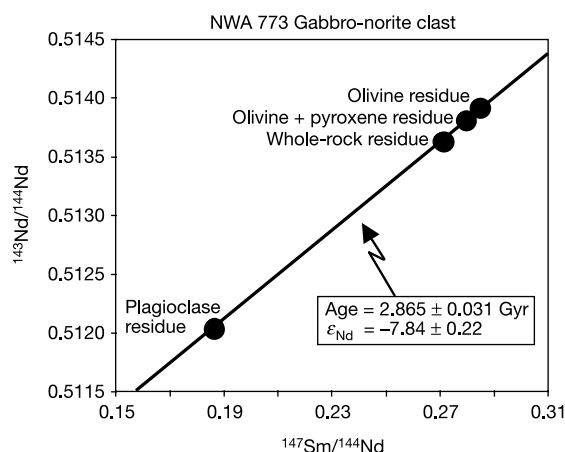


Figure 1 Sm–Nd isochron diagram for NWA 773, yielding the youngest lunar crystallization age of 2.865 ± 0.031 Gyr and the most negative ϵ_{Nd} value of -7.84 ± 0.22 .

Table 1 Rb–Sr and Sm–Nd isotopic data for NWA 773

Fraction	Whole-rock residue	Plagioclase residue	Olivine + pyroxene residue	Olivine residue
Weight fraction (mg)	8.28	8.33	28.38	37.80
Rb (p.p.m.)	0.266	2.79	0.130	0.127
Sr (p.p.m.)	8.719	101.09	2.810	2.711
$^{87}\text{Rb}/^{86}\text{Sr}^*$	0.08824 ± 44	0.07995 ± 40	0.13352 ± 67	0.13519 ± 68
$^{87}\text{Sr}/^{86}\text{Sr}^\dagger$	0.707849 ± 10	0.706845 ± 10	0.711716 ± 10	0.710068 ± 10
Sm (p.p.m.)	0.795	1.235	0.643	0.782
Nd (p.p.m.)	1.850	4.172	1.451	1.730
$^{147}\text{Sm}/^{144}\text{Nd}^\ddagger$	0.27085 ± 27	0.18661 ± 80	0.27922 ± 28	0.28494 ± 54
$^{143}\text{Nd}/^{144}\text{Nd}^\S$	0.513637 ± 10	0.512050 ± 10	0.513799 ± 10	0.513916 ± 11

NBS-987Sr standard ($N = 6$) = 0.710263 ± 16 , La Jolla Nd standard ($N = 8$) = 0.512873 ± 14 . Sr run on Re filaments with Ta_2O_5 . Nd run as NdO at 5×10^{-7} torr.

*Error limits apply to last digits and include a minimum uncertainty of 0.5% plus 50% of the blank correction for Rb and Sr added quadratically.

†Normalized to $^{86}\text{Sr}/^{86}\text{Sr} = 0.1194$. Uncertainties refer to last digits and are $2\sigma_m$, calculated from the measured isotopic ratios.

‡Error limits apply to last digits and include a minimum uncertainty of 0.1% plus 50% of the blank correction for Sm and Nd added quadratically.

§Normalized to $^{146}\text{Nd}/^{144}\text{Nd} = 0.7219$. Uncertainties refer to last digits and are $2\sigma_m$, calculated from the measured isotopic ratios.

region¹⁹. This suggests that the KREEP source may be significantly more evolved than has been previously estimated.

The initial $^{87}\text{Sr}/^{86}\text{Sr}$ ratio of NWA 773 is estimated to be 0.703568 ± 0.000032 from the plagioclase mineral fraction using the Sm–Nd age. The $^{87}\text{Rb}/^{86}\text{Sr}$ ratio of the NWA 773 source region is calculated to be 0.195, assuming an initial $^{87}\text{Sr}/^{86}\text{Sr}$ ratio of BABI (best achondrite basalt initial) at 4.558 Gyr ago. This value is significantly higher than the values calculated for other KREEP-rich rocks, which are typically less than 0.05 (refs 18, 20), confirming that NWA 773 is derived from a more evolved source.

The presence of KREEP-rich material in the lunar mantle has long been viewed as a potential heat source for melting of mafic cumulates^{4,6,7}. This stems from the observation that KREEP basalt has U and Th abundances that are enriched over chondrites by a factor of about 300 (ref. 4). The Rb/Sr ratios of the source regions of NWA 773 and other KREEP-rich samples have been calculated from their Rb–Sr isotopic systematics and plotted against age in Fig. 3. The Rb/Sr ratio of individual source regions is directly proportional to the amount of KREEP-rich material that they contain. For example, relatively KREEP-poor Apollo 15 glasses have low Rb/Sr ratios of about 0.014, whereas KREEP-rich mantle material has been estimated to have high Rb/Sr ratios of 0.11 (refs 4, 19, 21).

From Fig. 3 it is apparent that there is an inverse correlation between Rb/Sr ratio of the source region and the crystallization age. Samples with relatively old ages are derived from sources with low Rb/Sr ratios, whereas relatively young samples are derived from sources characterized by higher Rb/Sr ratios. Thus, the degree of

incompatible element enrichment inferred for the source regions of individual samples (that is, the amount of KREEP-rich material present) increases as the samples become younger. This correlation could reflect the need for the source regions of young magmas to contain a greater abundance of heat-producing elements to offset cooling associated with heat loss of the Moon through time. This correlation provides evidence that the presence of KREEP-rich material in the source region may indeed be required to initiate melting.

The correlation between the abundance of KREEP in the source and age also suggests that Mg-rich cumulates and KREEP-rich material must be physically associated with one another in the source region for both heat and mass to be exchanged between Mg-rich cumulates and KREEP-rich components. Therefore, melting-differentiation models that require the assimilation of KREEP-rich material in shallow-level magma chambers by Mg-rich magmas derived from deep in the mantle are less feasible. Instead, it seems that KREEP-rich material is mixed with Mg-rich cumulates in the source region.

Alternatively, Snyder *et al.*^{22,23} suggested that KREEP-rich samples from the Mg-suite and alkali suites were related by crystallization of closely related parental magmas. Several aspects of this model are consistent with the observation that young lunar samples have large KREEP geochemical signatures (Fig. 3). For example, the KREEP geochemical signature becomes more pronounced because the parental magma evolves through time. This, of course, requires KREEP-rich parental magmas to remain molten over the first 1.7 Gyr of lunar history. Although some thermal models predict KREEP-rich magmas to remain molten for very long periods of time^{5,7}, the Mg-rich nature of NWA773 is not consistent with this

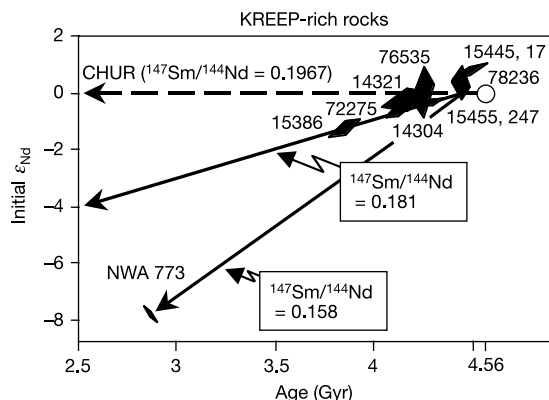


Figure 2 Age versus initial ϵ_{Nd} plot for KREEP-rich lunar rocks (Mg and alkali-suites, and KREEP basalts) demonstrating that NWA 773 is derived from the most evolved (LREE-enriched) lunar source region yet known. Data summary from refs 8 and 9. Solid lines are two-stage growth models for lunar source regions. The first stage represents growth in CHUR, between 4.558 Gyr ago and the time of magma ocean crystallization represented by the 4.42-Gyr KREEP model age. The second stage represents growth in the source regions of individual samples after 4.42 Gyr ago.

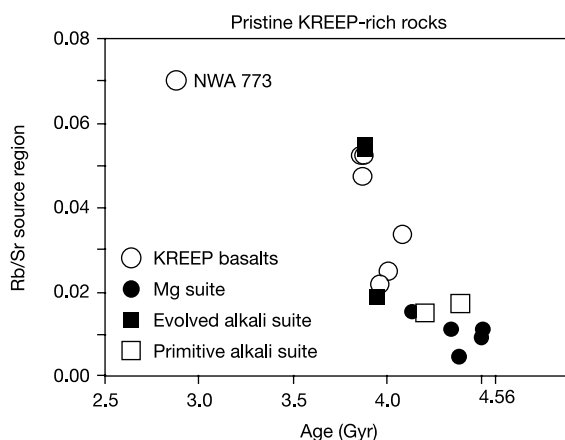


Figure 3 Plot of age versus Rb/Sr ratios of sample source regions calculated from their initial Sr isotopic compositions using a single-stage growth model. The inverse correlation suggests that KREEP is the heat source responsible for melting lunar samples.

hypothesis because it indicates that the parental magma is not particularly evolved.

There are two observations that are inconsistent with the hypothesis that KREEP initiates melting of mafic cumulates to produce young lunar magmas. First, there are no samples with ancient ages and large KREEP geochemical signatures in the sample collections. However, this could reflect a sampling bias introduced by the fact that all ancient KREEP-rich igneous samples are derived from two closely related plutonic suites²². The second observation is that radiometric ages determined on low- and high-Ti basalts that lack the KREEP geochemical signature are often relatively young (3.1 to 3.4 Gyr; refs 16, 24). This implies that another mechanism to melt sources with small amounts of KREEP-rich material is required. We speculate that KREEP-rich material is the heat source for these magmas as well. However, melting is not initiated at the site where this material resides; instead, heating of mafic cumulates by KREEP-rich materials promotes upwelling of diapirs in the mantle. Thus, melting could be initiated as pressure is released. If this mechanism is valid, it implies that KREEP-rich materials are directly or indirectly responsible for melting of the lunar mantle. □

Methods

The sample was washed and sonicated in four-times quartz-distilled water, followed by 0.5 M acetic acid. The sample was next crushed using a sapphire mortar and pestle, and sieved at 100–200 and 200–325 mesh. Mineral separations were begun using heavy liquids on both size fractions. Plagioclase floated in 2.85 g cm⁻³, whereas the mafic minerals sank. Hand-picking was used to separate pyroxene (brown) and olivine (green) mineral grains, as well as to purify the plagioclase fraction. Individual mineral fractions were then leached in 1 N HCl for 10 min in a sonicator before digestion. Chemical separations and isotope ratio measurements were done at the Radiogenic Isotope Laboratory, University of New Mexico, following standard silicate dissolution procedures, and involved cation chromatography using a combination of HCl and methalactic acids. Isotopic ratios were measured on a Micromass Sector 54 multi-collector thermal ionization mass spectrometer on Faraday cups in static mode. Rubidium, Sr, Sm, and Nd blanks measured during the course of the investigation averaged 7, 12, 7 and 8 pg, respectively. Normalization and standard values are given in Table 1.

Received 14 May; accepted 28 September 2004; doi:10.1038/nature03070.

- Wood, J. A., Dickey, J. S., Marvin, U. B. & Powell, B. N. Lunar anorthosites and a geophysical model of the Moon. *Proc. 1st Lunar Planet. Sci. Conf.* 965–988 (Pergamon, New York, 1970).
- Smith, J. A., et al. Petrologic history of the Moon inferred from petrography, mineralogy, and petrogenesis of Apollo 11 rocks. *Proc. 1st Lunar Planet. Sci. Conf.* 1149–1162 (Pergamon, New York, 1970).
- Snyder, G. A., Taylor, L. A. & Neal, C. R. A chemical model for generating the sources of mare basalts: Combined equilibrium and fractional crystallization of the lunar magmasphere. *Geochim. Cosmochim. Acta* **56**, 3809–3823 (1992).
- Warren, P. H. & Wasson, J. T. The origin of KREEP. *Rev. Geophys. Space Phys.* **17**, 73–88 (1979).
- Warren, P. H. The origin of pristine KREEP: Effects of mixing between urKREEP and the magmas parental to the Mg-rich cumulates. *Proc. 8th Lunar Planet. Sci. Conf.* 233–241 (1988).
- Hess, P. C. & Parmentier, E. M. Thermal evolution of a thicker KREEP liquid layer. *J. Geophys. Res.* **106**, 28023–28032 (2001).
- Wieczorek, M. A. & Phillips, R. J. The “Procellarum KREEP Terrane”: Implications for mare volcanism and lunar evolution. *J. Geophys. Res.* **105**, 20417–20430 (2000).
- Nyquist, L. E. & Shih, C.-Y. The isotopic record of lunar volcanism. *Geochim. Cosmochim. Acta* **56**, 2213–2234 (1992).
- Snyder, G. A., Borg, L. E., Nyquist, L. E. & Taylor, L. A. in *Chronology and Isotopic Constraints on Lunar Evolution* (eds Canup, R. & Righter, K.) 361–395 (Univ. Ariz. Press, Tucson, 2000).
- Fagan, T. J. et al. Northwest Africa 773: Lunar origin and iron-enrichment trend. *Meteorit. Planet. Sci.* **38**, 529–554 (2003).
- Jolliff, B. L., Korotev, R. L., Zeigler, R. A. & Floss, C. Northwest Africa 773: Lunar mare breccia with a shallow-formed olivine-cumulate component, very low Ti (VLT) heritage and a KREEP connection. *Geochim. Cosmochim. Acta* **67**, 4857–4879 (2003).
- Borg, L. E., Nyquist, L. E., Weismann, H., Shih, C.-Y. & Reese, Y. The age of Dar al Gani 476 and the differentiation history of the martian meteorites inferred from their radiogenic isotopic systematics. *Geochim. Cosmochim. Acta* **67**, 3519–3536 (2003).
- Ludwig, K. J. R. *Users Manual for Isoplot/Ex v. 2.49: A Geochronological Toolkit for Microsoft Excel* (Berkeley Geochronology Center Special Publication No. 1a, BGC, Berkeley, 2001) (<http://www.bgc.org/kpprogrammenu.html>).
- Fernandes, V. A., Burges, R. & Turner, G. ⁴⁰Ar–³⁹Ar chronology of lunar meteorites Northwest Africa 032 and 773. *Meteorit. Planet. Sci.* **38**, 555–564 (2003).
- Hiesinger, H., Head, J. W., Wolf, U., Jaumann, R. & Neukum, G. Ages and stratigraphy of mare basalts in Oceanus Procellarum, Mare Nubium, Mare Cognitum, and Mare Insularum. *J. Geophys. Res.* **108**, 1–27 (2003).
- Nyquist, L. E., Bansal, B. M., Wooden, J. L. & Wiesmann, H. Sr-isotopic constraints on the petrogenesis of Apollo 12 basalts. *Proc. 8th Lunar Planet. Sci. Conf.* 1383–1415 (Pergamon, New York, 1977).
- Papanastassiou, D. A. & Wasserburg, G. J. Rb–Sr age of troctolite 76535. *Proc. 7th Lunar Sci. Conf.* 2035–2054 (Pergamon, New York, 1976).

- Shih, C.-Y. et al. Age of pristine noritic clasts from lunar breccias 15445 and 15455. *Geochim. Cosmochim. Acta* **57**, 915–931 (1993).
- Warren, P. H. in *Workshop on Moon in Transition: Apollo 14, KREEP, and Evolved Lunar Rocks* 149–153 (LPI Technical Report 98–03, Lunar and Planetary Institute, Houston, 1989).
- Shih, C.-Y., Nyquist, L. E., Bansal, B. M. & Weismann, H. Rb–Sr and Sm–Nd chronology of an Apollo 17 KREEP basalt. *Earth Planet. Sci. Lett.* **108**, 203–215 (1992).
- Hughes, S. S., Delano, J. W. & Schmitt, R. A. Apollo 15 yellow-brown glass: Chemistry and petrogenetic relations to green volcanic glass and olivine-normative basalts. *Geochim. Cosmochim. Acta* **52**, 2379–2391 (1988).
- Snyder, G. A., Taylor, L. A. & Halliday, A. N. Chronology and petrogenesis of the lunar highlands alkali suite: Cumulates from KREEP basalt crystallization. *Geochim. Cosmochim. Acta* **59**, 1185–1203 (1995).
- Snyder, G. A., Neal, C. R., Taylor, L. A. & Halliday, A. N. Processes involved in the formation of magnesian-suite plutonic rocks from the highlands of the Earth's moon. *J. Geophys. Res.* **100**, 9365–9388 (1995).
- Dash, E. J. et al. Time of crystallization of a unique A15 basalt. *Lunar Planet. Sci. Conf.* **XX**, 218–219 (1989).

Acknowledgements We thank L. Nyquist for the use of his heavy liquid separations facility at the Johnson Space Center, M. Grady and S. Russell of the Natural History Museum, London for providing us with the sample, and two anonymous reviewers for their comments. This work was supported by NASA Cosmochemistry Program.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to L.E.B. (lborg@unm.edu).

Nonindependence of mammalian dental characters

Aapo T. Kangas, Alistair R. Evans, Irma Thesleff & Jukka Jernvall

Developmental Biology Program, Institute of Biotechnology, PO Box 56, FIN-00014, University of Helsinki, Helsinki, Finland

Studies of mammalian evolution frequently use data derived from the dentition^{1–4}. Dental characters are particularly central for inferring phylogenetic relationships of fossil taxa^{1–4}, of which teeth are often the only recovered part. The use of different aspects of dental morphology as phylogenetic signals implies the independence of dental characters from each other. Here we report, however, that, at least developmentally, most dental characters may be nonindependent. We investigated how three different levels of the cell signalling protein ectodysplasin (Eda)⁵ changed dental characters in mouse. We found that with increasing expression levels of this one gene, the number of cusps increases, cusp shapes and positions change, longitudinal crests form, and number of teeth increases. The consistent modification of characters related to lateral placement of cusps can be traced to a small difference in the formation of an early signalling centre at the onset of tooth crown formation. Our results suggest that most aspects of tooth shape have the developmental potential for correlated changes during evolution which may, if not taken into account, obscure phylogenetic history.

Although both new fossil and molecular data can be expected to resolve many of the phylogenetic incongruences^{6–10}, there is a continuing debate about the best way to use dental evidence in evolutionary taxonomy^{6,9,10}. The ‘total evidence’ view assumes that covarying characters manifest phylogenetic congruence, but other views assume that covariance of characters can be misinterpreted as a phylogenetic signal for functional or developmental reasons. As part of the same individual ontogeny, all characters obviously have the potential for nonindependent, or correlated, changes during evolution, but whether this could be the case for most mammalian dental characters used in evolutionary taxonomy remains unresolved. Features of the mammalian dentition, as part

of a repetitively structured system, have classically been suggested to share the same quantitative genetic underpinnings^{11–14}. Although an increasing number of heritable quantitative traits have been reported in teeth¹⁵, most identified gene mutations with dental effects are characterized by a total or partial lack of teeth and are thus uninformative for the character analysis of tooth morphology. Here we quantified effects of a relatively mild genetic change on tooth shape to test whether and how dental characters used in evolutionary taxonomy might be developmentally linked.

We analysed how tooth characters and their development were altered in two mutant mouse strains compared to the teeth of normal ‘wild type’ mice (*Mus musculus*). One was a natural mutant, *Tabby*, which lacks functional ectodysplasin, and the other was a transgenic mouse, K14-*Eda*, which has superfluous expression of *ectodysplasin* under keratin-14 promoter in the basal epithelial cell layer¹⁶. We considered the *ectodysplasin* expression levels of the three mouse strains as character states of one developmental variable, and examined how the morphology of teeth change when ectodysplasin activity levels rise from zero (*Tabby*) to normal (wild-type controls), and beyond normal (K14-*Eda*).

Lack of functional ectodysplasin in humans causes an X-linked hypohidrotic ectodermal dysplasia syndrome (X-HED) in which the development of ectodermal organs like teeth, hair and skin glands is defective¹⁷. The mouse equivalent⁵, *Tabby* mutant, was discovered more than 50 years ago and *Tabby* tooth morphology is characterized by missing cusps¹⁷. The presence of both human and mouse tooth phenotypes with reduced numbers of cusps and teeth due to *ectodysplasin* mutation, although not conclusive, is suggestive of the generality of the role of ectodysplasin in mammalian tooth development. Ectodysplasin, which is a member of the tumour necrosis factor (TNF) family of proteins, is a membrane protein which is cleaved to become freely diffusible in the extracellular space¹⁷, and has been implicated in modulating growth and differentiation of most epithelial organs^{17,18}. When the first morphologically distinguishable markers of developing teeth, dental placodes, appear, *ectodysplasin* expression is restricted to ectoderm that is adjacent to the tissues giving rise to teeth¹⁸. Ectodysplasin binds to its receptor Edar, which is expressed in dental placodes at

the initiation of tooth development and in the enamel knots during crown formation^{17,18}. The enamel knots, which express several signalling molecules, participate in regulating the formation of the tooth crown base and the individual cusps. The spatial patterns of enamel knots predict the species-specific cusp patterns¹⁹.

We used conventional light and high-resolution three-dimensional laser confocal microscopy^{20,21} to perform a character analysis. The results show that with increasing ectodysplasin signalling, many dental character states change beyond what is typically found within a species or even a genus (Fig. 1a). There is an overall increase in bucco-lingual distance between cusps. In *Tabby* the tips of the cusp pairs are typically fused, whereas in the wild-type mice the cusps are separated by an anteriorly oblique crest (Fig. 1a, b). In the K14-*Eda* mice, bucco-lingual separation of cusps has increased further with a straight transverse crest. Furthermore, in many teeth these transverse crests are connected by a longitudinal crest running along the midline of the crown (Fig. 1a, b). Longitudinal crests, not previously reported from mice or mouse mutants, are present in many other muroids and also in the Miocene members of the lineage leading to modern *Mus*²². The combination of straight lamellar-like transverse and central longitudinal crests is also typical of diprotodont marsupials (for example, kangaroos) and proboscideans (for example, African elephant). We note that first molars with a well-developed longitudinal crest connecting the protolophid to the hypolophid also have greater bucco-lingual cusp distances (Mann–Whitney *U* test, *P* = 0.025 to 0.035), suggesting that ectodysplasin signalling can cause the longitudinal crest formation indirectly, by affecting lateral placement of cusps.

A second distinct change correlating with the amount of ectodysplasin signalling is the number and shape of cusps. Compared to the anterior, trigonid part of the crown, the distal, talonid part is shallow in *Tabby* teeth (Fig. 1), and the distalmost cusp, the hypoconulid, is missing, whereas in the K14-*Eda* molars it has been incorporated into the wide posterior cingulid (Fig. 1). In addition to the missing hypoconulid, the characteristic anteroconid of the murid first lower molar is almost always missing in the *Tabby* mice. While these reductions in tooth crown features appear extreme, muroids such as the water rats (genus *Hydromys*), with a

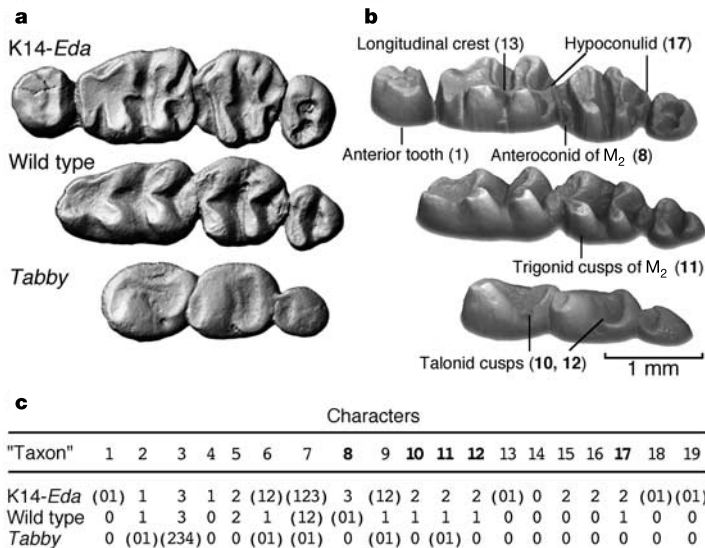


Figure 1 Mouse molar teeth differ in several characters in mice with no (*Tabby*), normal (wild type), and above normal (K14-*Eda*) *ectodysplasin* activity. Occlusal (a) and obliquely lingual (b) views show right lower molar rows. Buccolingual cusp pairs of *Tabby* are fused, while these form a transverse loph in K14-*Eda*. Several other character states differ among teeth (c), such as the presence of a tooth anterior to first molar (character 1) and

presence and shape of the hypoconulid (character 17). Many of the characters are polymorphic in both *Tabby* and K14-*Eda* mice (in parentheses). The non-polymorphic characters 10, 12 and 17, and also 8 and 11 with low degrees of polymorphism, are altered most consistently, owing to ectodysplasin signalling (shown in bold in b and c). For details, see Supplementary Information. M₂, lower molar.

primary diet consisting of fish and aquatic invertebrates²³, have molar morphologies reminiscent of *Tabby* molars. In contrast, the anteroconid of K14-*Eda* first molars have additional or better differentiated cusps. Limiting the size of the anteroconid anteriorly is an extra tooth in the K14-*Eda* mice in 58% of the tooth rows. This extra tooth is round and has one or two closely placed cusps. Its simple shape is reminiscent of the premolars of other rodent groups as well as premolars of the rodents implicated in the ancestry of muroids^{22,24}.

Our character state analysis shows that most of the *Tabby* and K14-*Eda* tooth features are polymorphic (Fig. 1c), indicating how an extreme drop or an increase in a gene activity may destabilize the developmental system. There is a tendency for more 'derived' character states to be present in the same tooth row but, in general, different parts of the crown have different sensitivities to changes in

ectodysplasin signalling. Character states of the talonid were the least polymorphic, and the first molar anteroconid and the third molar were the most polymorphic (Fig. 1c). These differences in character polymorphism, here resulting from a complete gene deactivation or multiplication, suggest that evolutionary changes stemming from small tinkering of gene expression level may produce, at least initially, stepwise changes in only one or a few characters. Whereas different molecules are likely to affect character covariance differently, ectodysplasin modifies characters 8, 10, 11, 12 and 17 most consistently (Fig. 1b, c). All these characters appear to measure how well individual cusp tips are differentiated and separated bucco-lingually, suggesting a role for ectodysplasin in the regulation of the iterative process of cusp formation along the tooth row¹⁹. We hypothesize that these characters would be the first visible effects of small changes in ectodysplasin levels.

Because our analyses show that several morphologically distinct characters can be altered as a function of changing expression level of a single gene, we next examined how the morphologies start to differ during development. We quantified the patterns of sonic hedgehog (*Shh*) expression from the onset of the first lower molar development. *Shh* is a signalling molecule required for the development of several organs and a marker for the overall signalling activity and size of the enamel knots¹⁹. In the enamel knots *Shh* expression allows the detection of lateral placement of the tooth cusps.

The results show that the earliest distinct differences in *Shh* expression domains are seen just before the formation of the first enamel knot (Fig. 2a). When the initial epithelial budding and first enamel knot appeared in wild-type mice, no *Shh* expression was detected in any of the studied *Tabby* tooth germs for two days. After this, *Shh* expression domains remain narrower relative to the wild-type molars throughout development and correspond in width to the fused cusps in fully formed teeth (Fig. 2a). In addition, the initiation of the talonid cusps of *Tabby* is delayed relative to the trigonid, resulting in the reduced talonid, analogous to the more primitive tribosphenic molar pattern. In contrast, in all K14-*Eda* mice studied, *Shh* expression continues uninterruptedly from the placode to the first enamel knot stage with initially a broader expression domain compared to the wild-type molars (Fig. 2a). The cusp-specific K14-*Eda* *Shh* expression of the following stage slightly lags behind that of the wild-type teeth, but widens faster as the teeth grow bigger (Fig. 2a). This delay in the widening of the K14-*Eda* first molar cusp spacing is concurrent with, and probably caused by, the expansion of the developing extra tooth anterior to it (Fig. 2b). These results indicate that modification of the forming primary enamel knot, from which the crown features develop, is enough to alter the states of several characters during the iterative process of cusp formation¹⁹. We interpret this to conform to the central role of morphodynamic²⁵ interplay between molecular signalling and growth, giving rise to the tooth shape. Thus, as shown here and supported by mathematical models²⁵, individual developmental variables can affect multiple characters simultaneously and each dental character does not necessarily require an individual 'genetic' code^{11–15,25}.

Although only about half of the K14-*Eda* tooth rows have an extra tooth, its development was initiated in all of the studied K14-*Eda* mice. However, the size of the *Shh* expression domain in the extra tooth is highly variable, suggesting that only tooth germs with the greatest initial knot activity develop into fully erupted extra teeth (Fig. 2b). This is indirectly supported by our discovery that wild-type mice have weak *Shh* expression activity in the same location at which the K14-*Eda* extra tooth develops (Fig. 2b). We note that although muroids have not developed lower premolars for over 45 million years^{22,24,26} it is intriguing that the wild-type mice have transient upregulation of *Shh* in the ancient premolar location. The incipient and continuing development of the extra tooth in wild-type mice and the K14-*Eda* mice, respectively, does not conflict

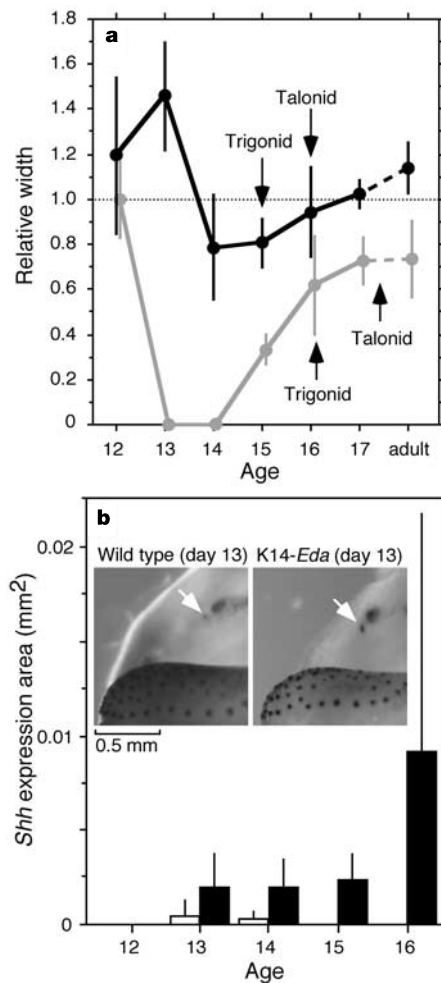


Figure 2 Dynamics of *Shh* expression in developing first molar (a) and extra tooth (b). Ratio-diagram of the width of *Shh* expression (cusps in adults) relative to wild-type molars (a) shows how the K14-*Eda* tooth expression domains (black line; error bars denote standard deviations) widen faster and surpass the wild type after day 17, while the *Tabby* tooth *Shh* expression domains (grey line) never reach the wild-type widths. *Shh* expression is not detected in *Tabby* from day 13 to 14 and in 40% of the wild-type teeth at day 13. *Shh* expression is continuous in K14-*Eda* teeth and relatively widest at day 12 and 13. All differences to the wild-type teeth are significant ($P < 0.05$, two-tailed test using 1,000 randomizations) except for *Tabby* at day 12 and K14-*Eda* at days 15 to 17. b, *Shh* is also upregulated (white arrows in the insets) anterior to the first enamel knot, transiently in wild-type mice (white bars) and prominently in the extra tooth of the K14-*Eda* mice (black bars). The developing taste papillae of the tongue express *Shh* (bottom of the insets).

with, and may support, the previous classification of minor epithelial invaginations anterior to the first molar as vestigial teeth²⁷. In wild ground squirrels (genus *Spermophilus*), Goodwin showed supernumerary teeth to be more prevalent in hybrids between two species²⁸, suggesting a finely tuned role for genetic background in normal suppression of extra tooth development.

In conclusion, our analyses show correlated changes in dental characters as a function of quantitative changes in intercellular signalling. In the cases where the signalling activity of ectodysplasin, or of another equivalent molecule, can be modulated exclusively in the teeth, through tooth-specific *cis*-regulation, for example, these results suggest that most aspects of tooth shape have the potential for correlated changes during evolution. Our analyses clearly do not exclude the potential for dental characters to change independently in evolution. On the contrary, the disparate sensitivities of different dental characters to changes in ectodysplasin signalling suggest that gradual changes in dental morphology are possible even due to changes in the activity of a single gene. However, particularly when intermediate fossil morphologies are scarce, developmental non-independence should not be excluded from the hypotheses considered in evolutionary taxonomy. Sample sizes allowing, alternative character-coding strategies can be recommended, particularly when characters related to lateral cusp placement are found to covary. Such a case is the simplified tooth morphology of *Tabby* mice which suggests that extreme reduction of dental features in mammalian lineages such as seals and whales may have required only a simple developmental change. This would not only have rendered some of the characters interdependent, but also probably increased the likelihood of convergent evolution. From the palaeo-ecological point of view, developmental linkage of dental characters may have both facilitated, and directed, a relatively rapid increase in number of cusps and crests during periods of substantial environmental change. □

Methods

Mouse strains and morphological analysis

A transgenic mouse line has been recently created which expresses *ectodysplasin* (splice form A1) under keratin-14 promoter (K14)¹⁶ in FVB/N mice (Jackson Laboratories). The *Tabby* allele was B6CBACa-A^w/A-Ta (stock no. JR 0314, Jackson Laboratories) and bred as described previously¹⁸. Wild-type mice were FVB/N and NMRI mice, which share similar character states. We analysed mice that lack or overexpress ectodysplasin (*Eda*) in preference to mice that lack or overexpress ectodysplasin receptor (*Edar*). There is no ectodysplasin signalling in either *Edar* null-mutants (*downless*) or *Tabby* mice (the *eda* null-mutant) and their teeth are morphologically similar. However, overexpressing ectodysplasin receptor under the K14 promoter will also change the pattern of cells responsive to ectodysplasin signalling, whereas the expression of *Edar* remains limited to enamel knots in the K14-*Eda* teeth. Thus, while K14-*Eda* mice can be considered as a character state change in only one developmental variable (signalling amount), K14-*Edar* mice represent character state changes both in strength of signalling and in distribution of responding cells. We note that K14-*Edar* tooth morphology is altered^{29,30}, but their morphology and pattern of variation differ from that of K14-*Eda* mice³⁰.

Here we analysed lower molars whose development is best understood. While upper molar morphology is also affected by ectodysplasin, only two of the K14-*Eda* mouse maxillae had an extra tooth. Thirty-six K14-*Eda*, 36 *Tabby*, and 34 wild-type (14 CBAT6T6 × NMRI and 20 FVB/N) tooth rows (both sides) were studied under a stereo microscope and photographed. Selected specimens were scanned using a laser confocal microscope at 8 μm xyz-resolution as described previously^{20,21} except that the teeth were first coated with eosin in thin paraloid solution (Rohm and Haas). Three-dimensional models (DEMs) were made using a three-dimensional version of the National Institutes of Health IMAGE (<http://www.physics.usyd.edu.au/physopt/3dview/>).

Tooth characters and character states

The states of 19 dental characters were evaluated for the three strains, largely following Meng *et al.*²⁴, with some additional characters and character states included to cover the range of variation. We note that the characters are not meant as an exhaustive analysis of the dental features. Rather, they are used as representative characters that would be readily coded for fossil taxa (see Supplementary Information).

Developmental analysis

Whole-mount *Shh* *in situ* hybridization was performed using InsituPro robot (Intavis), using the protocol described earlier¹⁶. The stained mandibles and teeth were photographed and *Shh*-expressing areas in the developing teeth were measured. The number of specimens measured for day 12, 13, 14, 15, 16 and 17 *Tabby* mice were 16, 16, 28, 13, 33 and 10, respectively. The corresponding number of specimens for K14-*Eda* mice were 29, 29,

23, 8, 25 and 8, and for wild-type teeth 30, 20, 15, 32, 12 and 12. We used FVB/N (K14-*Eda* controls) as the wild-type controls owing to the smaller differences between the wild-type teeth and K14-*Eda* teeth. We randomized the mouse strain assignments 1,000 times to test differences in mean expression domains.

Received 27 May; accepted 12 August 2004; doi:10.1038/nature02927.

1. Beard, K. C., Tong, Y. S., Dawson, M. R., Wang, J. W. & Huang, X. S. Earliest complete dentition of an anthropoid primate from the late middle Eocene of Shanxi Province, China. *Science* **272**, 82–85 (1996).
2. Flynn, J. J., Parrish, J. M., Rakotosamimanana, B., Simpson, W. F. & Wyss, A. R. A Middle Jurassic mammal from Madagascar. *Nature* **401**, 57–60 (1999).
3. Luo, Z. X., Cifelli, R. L. & Kielan-Jaworowska, Z. Dual origin of tribosphenic mammals. *Nature* **409**, 53–57 (2001).
4. Seiffert, E. R., Simons, E. L. & Attia, Y. Fossil evidence for an ancient divergence of lorises and galagos. *Nature* **422**, 421–424 (2003).
5. Srivastava, A. K. *et al.* The *Tabby* phenotype is caused by mutation in a mouse homologue of the *EDA* gene that reveals novel mouse and human exons and encodes a protein (ectodysplasin-A) with collagenous domains. *Proc. Natl Acad. Sci. USA* **94**, 13069–13074 (1997).
6. Naylor, G. J. P. & Adams, D. C. Are the fossil data really at odds with the molecular data? Morphological evidence for cetartiodactyla phylogeny reexamined. *Syst. Biol.* **50**, 444–453 (2001).
7. Woodburne, M. O., Rich, T. H. & Springer, M. S. The evolution of tribospheny and the antiquity of mammalian clades. *Mol. Phylogenet. Evol.* **28**, 360–385 (2003).
8. Archibald, J. D. Timing and biogeography of the eutherian radiation: fossils and molecules compared. *Mol. Phylogenet. Evol.* **28**, 350–359 (2003).
9. O'Leary, M. A., Gatesy, J. & Novacek, M. J. Are the dental data really at odds with the molecular data? Morphological evidence for whale phylogeny reexamined. *Syst. Biol.* **52**, 853–864 (2003).
10. Geisler, J. H. & Uhen, M. D. Morphological support for a close relationship between hippos and whales. *J. Vert. Paleontol.* **23**, 991–996 (2003).
11. Bateson, W. *Materials for the Study of Variation, Treated with Special Regard to Discontinuity in the Origin of Species* (Macmillan, London, 1894).
12. Butler, P. M. The ontogeny of molar pattern. *Biol. Rev.* **31**, 30–70 (1956).
13. Van Valen, L. An analysis of developmental fields. *Dev. Biol.* **23**, 456–477 (1970).
14. Weiss, K. M., Stock, D. W. & Zhao, Z. Dynamic integrations and the evolutionary genetics of dental patterning. *Crit. Rev. Oral Biol. Med.* **9**, 369–398 (1998).
15. Hlusko, L. J. Integrating the genotype and phenotype in hominid paleontology. *Proc. Natl Acad. Sci. USA* **101**, 2653–2657 (2004).
16. Mustonen, T. *et al.* Stimulation of ectodermal organ development by ectodysplasin-A1. *Dev. Biol.* **259**, 123–136 (2003).
17. Mikkola, M. & Thesleff, I. Ectodysplasin signaling in development. *Cytokine Growth Factor Rev.* **14**, 211–224 (2003).
18. Laurikkala, J. *et al.* TNF signaling via the ligand-receptor pair ectodysplasin and edar controls the function of epithelial signaling centers and is regulated by Wnt and activin during tooth organogenesis. *Dev. Biol.* **229**, 443–455 (2001).
19. Jernvall, J., Keränen, S. V. E. & Thesleff, I. Evolutionary modification of development in mammalian teeth: Quantifying gene expression patterns and topography. *Proc. Natl Acad. Sci. USA* **97**, 14444–14448 (2000).
20. Jernvall, J. & Selänne, L. Laser confocal microscopy and geographic information systems in the study of dental morphology. *Paleo. Electronica* **2**, 1–18 (http://www-odp.tamu.edu/paleo/1999_1/confocal/issue1_99.htm) (1999).
21. Evans, A. R., Harper, I. S. & Sanson, G. D. Confocal imaging, visualization and 3-D surface measurement of small mammalian teeth. *J. Microsc.* **204**, 108–118 (2001).
22. Flynn, J. J., Jacobs, L. L. & Lindsay, E. H. in *Evolutionary Relationships among Rodents. A Multidisciplinary Analysis* (eds Luckett, W. P. & Hartenberger, J.-L.) 589–616 (NATO ASI Series, Plenum, New York, 1985).
23. Nowak, R. M. *Walker's Mammals of the World*, 5th edn (Johns Hopkins Univ. Press, Baltimore, 1991).
24. Meng, J., Hu, Y. M. & Li, C. K. The osteology of Rhombomylus (mammalia, glires): Implications for phylogeny and evolution of glires. *Bull. Am. Mus. Nat. Hist.* **275**, 1–247 (2003).
25. Salazar-Ciudad, I. & Jernvall, J. A gene network model accounting for development and evolution of mammalian teeth. *Proc. Natl Acad. Sci. USA* **99**, 8116–8120 (2002).
26. Dawson, M. R. A. & Tong, Y. New material of Pappocricetodon schaubi, an Eocene rodent (Mammalia: Cricetidae) from the Yuanqu basin, Shanxi province, China. *Bull. Carnegie Mus. Nat. Hist.* **34**, 278–285 (1998).
27. Viriot, L., Peterkova, R., Peterka, M. & Lesot, H. Evolutionary implications of the occurrence of two vestigial tooth germs during early odontogenesis in the mouse lower jaw. *Connect. Tiss. Res.* **43**, 129–133 (2002).
28. Goodwin, H. T. Supernumerary teeth in Pleistocene, recent and hybrid individuals of the *Spermophilus richardsonii* complex (Sciuridae). *J. Mamm.* **79**, 1161–1169 (1998).
29. Tucker, A. S., Headon, D. J., Courtney, J. M., Overbeek, P. & Sharpe, P. T. The activation level of the TNF family receptor, Edar, determines cusp number and tooth number during tooth development. *Dev. Biol.* **268**, 185–194 (2004).
30. Pispas, J. *et al.* Tooth patterning and enamel formation can be manipulated by misexpression of TNF receptor Edar. *Dev. Dyn.* **231**, 433–441 (2004).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank F. Ankel-Simons, M. Fortelius, J. Eronen, K. Kavanagh, S. King, J. Meng, P. Munne, T. Mustonen, J. Pispas, S. Pochron, I. Salazar-Ciudad, E. Seiffert, and P. C. Wright for comments or help on this work, which was supported by the Academy of Finland.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to J.J. (jernvall@fastmail.fm).

Disposable-soma senescence mediated by sexual selection in an ungulate

Juan Carranza, Susana Alarcos, Cristina B. Sánchez-Prieto, Juliana Valencia & Concha Mateos

Biology and Ethology Unit, Universidad de Extremadura, 10071 Cáceres, Spain

Senescence may result from an optimal balance between current reproductive investment and bodily repair processes required for future reproduction¹, a theoretical prediction difficult to prove especially in large, long-lived animals. Here we propose that teeth that have fixed dimensions early in life, but that wear during chewing, can be taken as a measure of total lifetime 'repair', and their wear rate as a measure of current expenditure in performance. Our approach also considers the sexual selection process to investigate the advance of senescence in males compared with females, when selection favouring competition over mates reduces the reproductive lifespan of males². We studied carcasses of 2,141 male and 739 female red deer (*Cervus elaphus*) of different ages, finding that male molariform teeth emerged at a far smaller size than expected from body size dimorphism. This led to higher workload, steeper wear rate and earlier depletion of male teeth than in females, in concordance with sex-specific patterns of lifetime performance and reproduction. These findings provide the empirical support for the disposable-soma hypothesis of senescence³, which predicts that investment in bodily repair will decrease when the return from this investment may not be realized as a result of other causes that limit survival or reproduction.

Senescence is defined as the progressive loss of function accompanied by decreased survival and reproductive rate with increasing age. Three main hypotheses have been proposed to explain how senescence may evolve: the mutation-accumulation hypothesis^{4,5}, the antagonistic pleiotropy hypothesis^{6–8} and the disposable soma hypothesis^{1,3,9}. In most cases, empirical data can hardly differentiate between them. For all three hypotheses, the evolutionary basis of senescence is the weakening of selection against the loss of function at older ages, because of lower survival or reproductive chances caused by extrinsic factors^{7,10}. Experiments have been performed with short-lived organisms such as *Drosophila* and *Caenorhabditis elegans* to modify the extrinsic causes of survival or reproduction, with effects on the onset of senescence⁸. The experimental approach is less appropriate for investigating the causes of senescence in long-lived animals, for which comparison between populations may be an alternative. For birds and mammals, comparisons between populations revealed a relationship between extrinsic mortality before senescence and the rate of senescence¹¹. However, the relationship between extrinsic mortality and senescence is common to all evolutionary explanations and cannot differentiate between hypotheses.

In studying the disposable-soma hypothesis, measuring body repair entails practical difficulties that hamper empirical support for the hypothesis. In red deer, as in most mammals, once the permanent teeth have erupted, the only changes to dentition are the gradual loss of teeth through wear. Although ungulate molars internally lay down secondary dentine, they cannot be repaired externally once worn. In many ungulates tooth wear has been shown to be a proximal cause of senescence^{12–16}. A corollary of the disposable-soma hypothesis for teeth is therefore that they are produced with properties that allow the organism to use them during its expected reproductive lifespan.

We proposed that although selection in male red deer produced

an increase in body size in comparison with females, it should have promoted only a slight increase in the size of teeth, because under the same selective process reproductive lifespan was reduced relative to that of females^{2,17}.

Maximum longevity recorded in our sample was 18 years for females and 13 years for males. On the basis of age distribution (Table 1), survival from age 2 years onwards followed log-log functions for males and females (male survival = $\exp(-\exp(0.76 - 0.22 \times \text{age}))$, $r = 0.987$, $N = 11$, $P < 0.0001$; female survival = $\exp(-\exp(0.85 - 0.15 \times \text{age}))$; $r = 0.974$, $N = 15$, $P < 0.0001$; see ref. 18). Accordingly, age-specific mortality rates fitted quadratic concave functions that were more pronounced in males than in females (for males, $y = 0.61 - 0.10x + 0.01x^2$, $r = 0.831$, $N = 12$, $P < 0.005$; for females, $y = 0.33 - 0.05x + 0.01x^2$, $r = 0.920$, $N = 12$, $P < 0.0002$; see also refs 17, 19, 20).

Plots of body mass against age followed quadratic convex functions with maximum values at 6.9 years in males and at 11.8 years in females (Fig. 1a). The relationship between crown height of the mandibular first molar (M1) and age followed quadratic concave functions in males and females (Fig. 1b). Crown height and age-specific rates of tooth wear differed between the sexes (analysis of covariance for natural logarithm of crown height with age as a covariate: for effect of sex factor, $F = 27.22$, d.f. 1, 2611, $P < 0.0001$; for interaction between sex and age, $F = 100.25$, d.f. 1, 2611, $P < 0.0001$; see also ref. 16). Even though males at 2 years of age started with slightly higher crowned first molars (11.0 mm compared with 10.8 mm in females), they were depleted at a higher rate (average 1.08 mm per year between 2 and 12 years of age) than in females (average 0.62 mm per year between 2 and 18 years of age) and, as a consequence, the absolute crown height of males (that is,

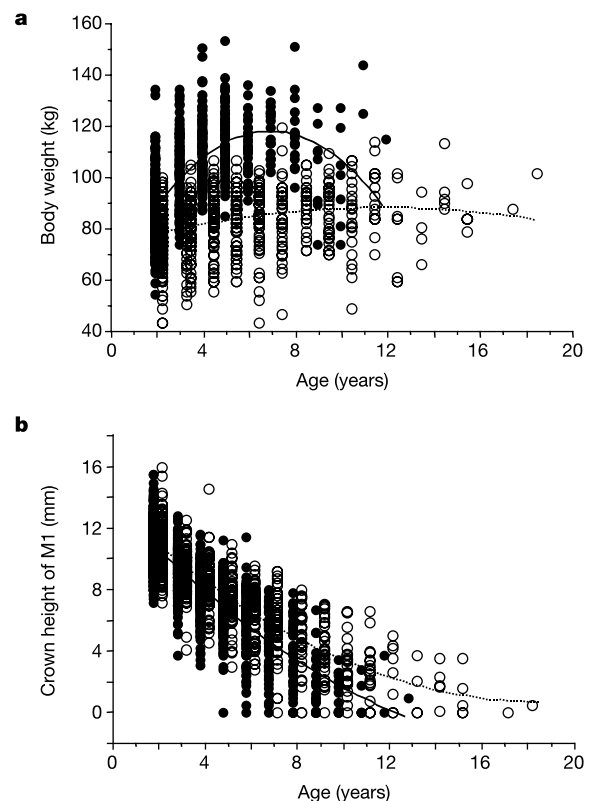


Figure 1 Performance and tooth wear throughout life in males and females. **a**, Variation in body mass with age for males (filled circles) and females (open circles). **b**, Crown height of mandibular first molar (M1) with respect to age for males (filled circles) and females (open circles). For details of functions see Supplementary Information.

Table 1 Frequency distribution of males and females of different ages used in the study

Sex	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Female	196	125	93	69	66	50	36	36	25	19	8	4	5	5	0	1	1
Male	879	433	293	232	125	70	48	34	17	7	2	1					

The total number of males was 2,141; the total number of females was 739.

even without adjusting for differences in body size) was lower than that of females from 3 years of age onwards.

Standardized sexual dimorphism for variables denoting tooth size ranged from 0.87 to 1.12. By contrast, standardized sexual dimorphism for variables related to body size and body mass ranged from 1.16 to 1.48 (Fig. 2). The crown height of the second molar (M2) was a striking extreme case, even showing a slightly smaller size in males than in females. It should be noted that the low dimorphism among tooth dimensions was not a consequence of previous wear, because for structures that did not increase with age we used the measurements at 2 years of age (see Methods and Supplementary Information) when previous wear had been minimal or even non-existent for M2. Mean dimorphism for the five tooth dimensions (occlusal surface area (OSA) variables are derived from them) was significantly lower than body size dimorphism (1.04 ± 0.10 s.d. compared with 1.34 ± 0.11 s.d.; Mann–Whitney U : $z = 2.842$, $N_1 = 5$, $N_2 = 7$, $P = 0.005$, which did not depend on the extreme case; $P = 0.008$ after excluding M2). Dimorphism in all tooth dimensions lay below expectations from either isometric²¹ or negative allometric²² scaling with body size (Fig. 2). Although mandible size maintained an isometric degree of dimorphism, the postcanine tooth row (PCTR) did not match the expected increment in size (Fig. 2).

As molariform teeth wear, they experience changes that affect the working surface area. Particularly because of the mesio-distal flaring out of molars, wear affects the width of the crown very little but reduces its length. Consequently, M1 OSA decreased slightly with age. The length of the whole PCTR as well as the total OSA followed convex patterns but varied little with age (see Supplementary Information).

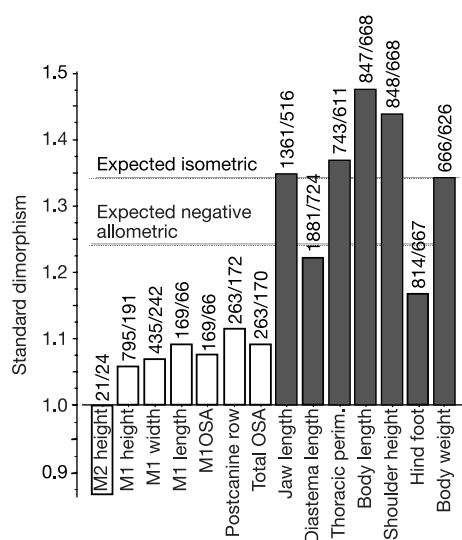


Figure 2 Sexual dimorphism (male/female ratio, standardized; see Methods), for different dental (open bars) and body size (filled bars) traits. Sample sizes (males/females) are indicated over the bars. Predicted standardized dimorphisms are shown, based on isometric (proportional to mass) and negative allometric (mass^{0.75}) assumptions.

According to current scaling theory^{21,22}, postcanine working surfaces should scale with (body mass)^{0.66} (isometric assumption for surfaces²¹) or (body mass)^{0.5} (negative allometric assumption for scaling surfaces with mass predicted by the mechanics of chewing²²). We estimated individual workloads on postcanine working surfaces, dividing body mass taken to both scaling powers (0.66 and 0.5) by M1 OSA and total OSA (see Supplementary Information). In all age groups, male teeth supported a higher workload than female teeth (effect of sex in two-factor analyses of variance: for isometric assumption, M1 OSA: $F = 44.187$, d.f. 1, 348, $P < 0.0001$; total OSA: $F = 29.913$, d.f. 1, 207, $P < 0.0001$; for negative allometric assumption, M1 OSA: $F = 24.978$, d.f. 1, 348, $P < 0.0001$; total OSA: $F = 15.988$, d.f. 1, 207, $P < 0.0001$).

To maintain the same workload, the male/female ratio of dental working surfaces should match the male/female ratio of body mass taken to the scaling powers. Both the ratio of M1 OSA and that of total OSA were far below expectations, indicating that postcanine

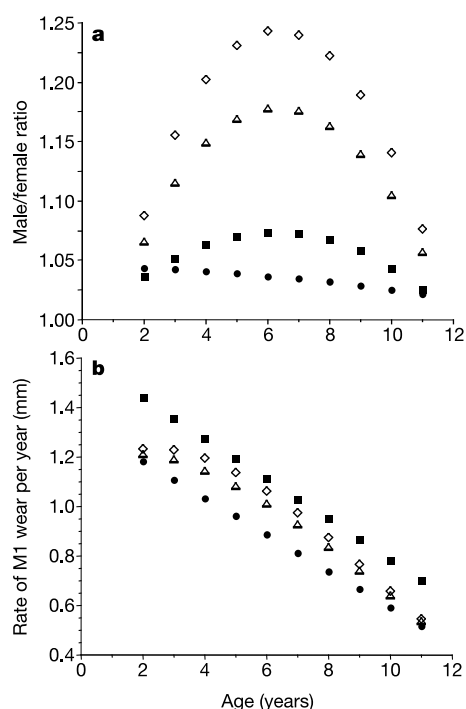


Figure 3 Postcanine working areas and wear. **a**, Expected male/female ratios for postcanine occlusal surface areas (OSA) according to isometric (exponent 0.66 for surfaces; open diamonds) and negative allometric (exponent 0.5, according to fracture scaling; open triangles) assumptions, and observed male/female ratios of total OSA (filled squares) and M1 OSA (filled circles) with age. **b**, Observed male (filled squares) and female (filled circles) wear rates, and predicted male wear rates based on both isometric (open diamonds) and negative allometric (open triangles) scaling criteria. These predictions were obtained by multiplying female rates by the ratio of expected to observed curves in **a** (using the observed M1 OSA because we wished to predict M1 wear).

occlusal surfaces of males were insufficient for the requirements derived from an increased body size in comparison with females (Fig. 3a).

To obtain a prediction of expected male M1 wear rate based on M1 relative size with respect to females, we corrected the pattern of M1 wear rate for females according to the ratio of M1 OSA relative to the expected ratio under both scaling criteria. Predicted male M1 wear rates were higher than female rates but did not reach the actual rates observed for males (Fig. 3b). This indicates that males are processing disproportionately more food than females, or food of lower quality^{19,23–26}, but also indicates that the size of male teeth should be positively allometric with respect to females to maintain durability. Predicted wear rates were closer to male observed values for prime ages (about 5–8 years). At younger ages, males depleted their dentine at a very high rate. A likely cause is the strategy of mass gain of males. Younger males experience a very high rate of body growth in comparison with females, and the difference between observed and expected wear was related positively to annual increment of body mass (from 2 to 6 years: $r = 0.940$, $N = 5$, $P = 0.017$; and $r = 0.950$, $N = 5$, $P = 0.013$, under both criteria respectively).

Sexual selection has moved red deer males towards a more semelparous schedule (about 6 years of reproductive lifespan) compared with females (about 14 years (refs 17, 19)). Our results clearly show that tooth size in males has been produced by selection to reach only the age above which reproductive chances are very low for causes other than tooth wear, which is strong evidence for planned senescence. Two main theories can account for the evolution of planned senescence: antagonistic pleiotropy⁶ and disposable soma⁹. Distinction between them is frequently blurred because both concepts are founded on the idea of trade-offs between benefits in early and late life¹. Antagonistic pleiotropy suggests that genes conferring positive benefits in early life may be favoured by selection even if they entail reproductive costs in later life. By contrast, disposable soma proposes that the trade-off between bodily use and repair is expected to lead to limited investments in the durability and repair of somatic structures. Therefore, although departures from optimal investment must also entail costs, the idea of early benefits and late costs is more fundamental to antagonistic pleiotropy, whereas planned durability is more compatible with disposable soma, thus providing a way of distinguishing between the hypotheses.

Possible early benefits of producing smaller teeth are to save materials and to accelerate tooth eruption. For materials, quantities seem too small to contribute significantly to skeleton or antler formation. For tooth eruption, M1 is already functional in both males and females at 1 year of age, and weaning is delayed in male calves¹⁹. In our data, although M1 thickness was smaller in males from 3 years onwards, the stained part of the M1 crown (see Methods) was higher in males until 4 years of age, indicating some delay in eruption compared with females. For possible costs in late life, depleted teeth are related to lower feeding efficiency^{12,14–16,20}, but this takes place in male red deer by the age of 11–12 years. Body performance starts to decline well before this age (Fig. 1), and probably also reproductive success, as indicated by data from other populations, once corrected for differences in longevity^{17,19}. The smaller size of male M2 also supports disposable-soma predictions, because it erupted 1 year later than M1 and has to be used during an even shorter lifespan. Therefore, in the situation where males have evolved a significantly shorter reproductive lifespan arising from intrasexual competition over mates^{2,17,19}, the disposable-soma hypothesis^{1,3,9} seems to be more appropriate for providing a ready explanation of why molars in this sex are under-provisioned for long-term function in comparison with females. Nevertheless, evidence in favour of the disposable-soma hypothesis is not opposed to antagonistic pleiotropy, but simply highlights how the mechanisms of somatic maintenance and repair have a role in the

evolution of trade-off strategies underlying senescence¹.

We suggest that the comparison between sexes of dimorphic species for which the sexual selection process is well known opens up an interesting avenue of research, in which functional and proximal causes of the differences in reproductive patterns and senescence may be better understood than when comparing between species. In this context, the study of the features and wear rates of permanent teeth, or any other permanent structures, provide a promising field of evidence to explore how selection favours senescence. □

Methods

Red deer populations

Animals were 2,141 males and 739 females from 2 to 18 years of age, harvested in hunts in natural populations of Iberian red deer (*Cervus elaphus hispanicus*) in southwestern Spain. Thousands of stags are hunted every year in Spanish commercial hunts. Normally, every male deer aged 2 years or more can legally be shot. For females there are other non-commercial, management hunts aimed at reducing density, in which any female of any age can be culled, including calves. Hunting pressure on males is basically regulated by allowing only one hunting action per year in the same area and a minimum distance between hunters, whereas for females annual quotas are authorized by local government. In no instance did our study lead to the shooting of additional deer (for examples of use of harvesting data, see refs 16, 20, 27).

Recording of field data

Each hunting day we visited the place in the field where culled animals were gathered. From each animal we recorded some measurements in the field, and removed mandibles for further laboratory measurements. Field measurements were as follows: body mass of the complete animal (measured with electronic scales, 1–3 h after death, to the nearest 0.5 kg), body length (from nose to tip of tail excluding hair, following the dorsal body contour), shoulder height (from the vertebral column between the shoulders, the lateral contour along one foreleg, excluding the hoof), hind foot (straight length from the rear, outermost point of the hock to the beginning of the hoof) and thoracic perimeter (around the trunk behind the forelegs).

It was not possible to take all measurements for all individuals in the field, so numbers of samples for some measurements are smaller than the total number of animals recorded and are variable between measurements. For some hunting actions we had access to only the heads, not the whole bodies, which produced differences between sample sizes for head and body measurements.

Recording of laboratory data

Mandibles were used in estimating age and for taking measurements of jawbone and dental pieces, including wear. Age was estimated by counting cementum growth marks at the interradicular pad under the first molar²⁸, and checked by eruption patterns in younger animals. Ages are expressed in completed years from birth, so an animal aged n years is living its $n + 1$ year of life, as used for humans.

Tooth wear was estimated from the crown height of the first molar (M1), which was measured, with a calliper, by transverse cross-section of the M1 between its mesial and distal halves, as the distance in millimetres from the lowest point at mid-crown down to the central peak where the dentine touches the cementum layers. In some cases, the stained part of M1 crown height was measured from the disto-buccal cusp to the enamel/cementum line. For a subsample of individuals, we also measured crown height in the second molar (M2) as in M1.

From a sample of mandibles we took the following further measurements: jawbone length (the maximum straight line from the rearmost point of the ramus to the most distal part of the incisor alveoli excluding the incisors), diastema length (straight line between the lower incisor-shaped canine and the first premolar), PCTR (length of the PCTR of one ramus of the jaw (premolar + molar row) in individuals with the third molar erupted), M1 width (bucco-lingual diameter of the M1 occlusal plane), M1 length (mesio-distal diameter of the M1 occlusal plane), M1 OSA (width × length of M1 occlusal plane) and total OSA (estimated as the product of PCTR and M1 width).

Data analyses

The proportion of survivors at age i was computed for individuals of age 2 years as the number of individuals of age i divided by the number of individuals of age 2. Age-specific mortality for age i was computed as $1 - [\text{survivors at age}(i + 1)/\text{survivors at age } i]$. Because mortality was estimated from transverse data, some cases of age-specific mortality of 0 or less occurred and were eliminated from the analysis because they have no biological meaning. Biased hunting might have affected only slightly our estimation of survival. Although some increase in mortality might be expected at age 2 years in males because hunting is allowed on stags from this age, our own experience with game managers indicates that culling intensity is not likely to vary for ages greater than 2 years. We can therefore safely assume that age-specific relative mortality at least for mature and older males is not dependent on hunting pressure, and the same applies to the maximum age recorded (which conforms to tooth wear data). The assumption of age-independent culling is safe for females, because they are culled simply to reduce density without any selective criteria.

Sexual dimorphism was measured as the size of a morphometric variable in males divided by that in females. To calculate dimorphism, we first fitted the relationship

between trait values and age for males and females, and used maximum values of the curves for each sex. In particular, when relationships were decreasing linear or concave quadratic, we used mean values at 2 years of age; for convex quadratic relationships we used maximum values for male and female curves, and for asymptotic relationships we used the asymptotes.

Dimorphisms of linear measurements were standardized by elevating them to the third power and surfaces were multiplied by their square root, to make them comparable with volumes and weights. Dependent variables were explored for normality and homogeneity of variances. Crown height was transformed to $\ln(\text{crown height} + 1)$ to obtain linear relationships with age. The significance level was set at $P = 0.05$ and all P values are two-tailed.

Received 8 July; accepted 8 September 2004; doi:10.1038/nature03004.

- Kirkwood, T. B. L. & Austad, S. N. Why do we age? *Nature* **408**, 233–238 (2000).
- Andersson, M. *Sexual Selection* (Princeton Univ. Press, Princeton, New Jersey, 1994).
- Kirkwood, T. B. L. in *Handbook of the Biology of Ageing* (eds Finch, C. E. & Schneider, E. L.) 27–44 (Van Nostrand Reinhold, New York, 1985).
- Medawar, P. B. Old age and natural death. *Mod. Q.* **1**, 30–56 (1946).
- Edney, R. B. & Gill, R. W. Evolution of senescence and specific longevity. *Nature* **220**, 281–282 (1968).
- Williams, G. C. Pleiotropy, natural selection and the evolution of senescence. *Evolution* **11**, 398–411 (1957).
- Hamilton, W. D. The moulding of senescence by natural selection. *J. Theor. Biol.* **12**, 12–45 (1966).
- Rose, M. R. *Evolutionary Biology of Aging* (Oxford Univ. Press, New York, 1991).
- Kirkwood, T. B. L. & Rose, M. R. Evolution of senescence: late survival sacrificed for reproduction. *Phil. Trans. R. Soc. Lond. B* **332**, 15–24 (1991).
- Haldane, J. B. S. *New Paths in Genetics* (Allen & Unwin, London, 1941).
- Ricklefs, R. E. Evolutionary theories of ageing: confirmation of a fundamental prediction, with implications for the genetic basis and evolution of lifespan. *Am. Nat.* **152**, 24–44 (1998).
- Tyler, N. J. C. *Natural Limitation of the Abundance of the High Arctic Svalbard Reindeer*. Thesis, Univ. Cambridge (1987).
- Skogland, T. Tooth wear by food limitation and its life history consequences in wild reindeer. *Oikos* **51**, 238–242 (1988).
- Gaillard, J.-M., Delorme, D., Boutin, J.-M., Van Laere, G. & Pradel, R. Roe deer survival patterns: a comparative analysis of contrasting populations. *J. Anim. Ecol.* **62**, 778–791 (1993).
- Ericsson, G. & Wallin, K. Age-specific moose (*Alces alces*) mortality in a predator-free environment: evidence for senescence in females. *Ecoscience* **8**, 157–163 (2001).
- Loe, L. E., Mysterud, A., Langvatn, R. & Stenseth, N. C. Decelerating and sex-dependent tooth wear in Norwegian red deer. *Oecologia* **135**, 346–353 (2003).
- Clutton-Brock, T. H., Albon, S. D. & Guinness, F. E. in *Reproductive Success* (ed. Clutton-Brock, T. H.) 325–343 (Chicago Univ. Press, Chicago, 1988).
- Gompertz, B. On the nature of the function expressive of the law of human mortality and on the new mode of determining life contingencies. *Phil. Trans. R. Soc. Lond.* **115**, 513–585 (1825).
- Clutton-Brock, T. H., Guinness, F. E. & Albon, S. D. *Red Deer. Behaviour and Ecology of Two Sexes* (Edinburgh Univ. Press, Edinburgh, 1982).
- Mysterud, A., Yoccoz, N. G., Stenseth, N. C. & Langvatn, R. Effects of age, sex, and density on body weight of Norwegian red deer: evidence of density-dependence senescence. *Proc. R. Soc. Lond. B* **268**, 911–919 (2001).
- Fortelius, M. Ungulate cheek teeth: developmental, functional, and evolutionary interrelations. *Acta Zool. Fenn.* **180**, 1–76 (1985).
- Lucas, P. W. *Dental Functional Morphology* (Cambridge Univ. Press, Cambridge, 2004).
- Staines, B. W. & Crisp, J. M. Observations on food quality in Scottish red deer (*Cervus elaphus*) as determined by chemical analysis of the rumen contents. *J. Zool.* **185**, 253–259 (1978).
- Clutton-Brock, T. H., Iason, G. R. & Guinness, F. E. Sexual segregation and density related changes in habitat use in female and male red deer (*Cervus elaphus* L.). *J. Zool.* **211**, 275–289 (1987).
- Mysterud, A. The relationship between ecological segregation and sexual body-size dimorphism in large herbivores. *Oecologia* **124**, 40–54 (2000).
- Conradt, L., Gordon, I. J., Clutton-Brock, T. H., Thomson, D. & Guinness, F. E. Could the indirect competition hypothesis explain inter-sexual site segregation in red deer (*Cervus elaphus* L.)? *J. Zool.* **254**, 285–293 (2001).
- Bonenfant, C., Gaillard, J.-M., Loison, A. & Klein, F. Sex-ratio variation and reproductive costs in relation to density in a forest-dwelling population of red deer. *Behav. Ecol.* **14**, 862–869 (2003).
- Mitchell, B. Growth layers in dental cement for determining the age of red deer (*Cervus elaphus* L.). *J. Anim. Ecol.* **36**, 279–293 (1967).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank T. Kirkwood, P. W. Lucas, A. Mysterud and J. Pérez-Barbería for comments; R. Álvarez, L. Castillo, P. Cidoncha, A. Flores, B. Gutiérrez, J. G. Martínez, Y. Moreno, S. del Río and B. Sánchez for help in field and laboratory work, the Dirección General de Medio Ambiente de Extremadura for permissions and facilities, and J. A. Campón for allowing data and sample collection. Financial support came from Spanish Ministry of Science, FEDER and Junta de Extremadura (Consejería de Agricultura y Medio Ambiente, and Consejería de Educación Ciencia y Tecnología). C.B.S.P. and S.A. were supported by a predoctoral grant of the Autonomic Government of Extremadura.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to J.C. (carrazna@unex.es).

A socially enforced signal of quality in a paper wasp

Elizabeth A. Tibbetts^{1*} & James Dale²

¹Department of Neurobiology and Behavior, Cornell University, Ithaca, New York 14853, USA and Center for Insect Sciences & Division of Neurobiology, University of Arizona, Tucson, Arizona 85721, USA

²Department of Biological Sciences, Simon Fraser University, Burnaby, British Columbia V5A 1S6, Canada

* Present address: Center for Insect Sciences & Division of Neurobiology, University of Arizona, Tucson, Arizona 85721, USA

Organisms use signals of quality to communicate information about aspects of their relative phenotypic and genetic constitution^{1–4}. Badges of status^{5–7} are a subset of signals of quality that reveal information about an individual's size and dominance. In general, signals of quality require high and differential costs to remain honest^{1,2} (that is, prevent low-quality cheaters from exploiting any fitness benefits associated with communicating high quality). The theoretically required costs for badges of status remain controversial because the development (or 'production') of such signals often seems to be relatively cost-free^{5,6,8}. One important hypothesis is that such signals impose social (or 'maintenance') costs incurred through repeated agonistic interactions with other individuals^{9–12}. However, convincing empirical evidence for social costs remains elusive^{6,7}. Here we report social costs in a previously undescribed badge of status: the highly variable black facial patterns of female paper wasps, *Polistes dominulus*. Facial patterns strongly predict body size and social dominance. Moreover, in staged contests between pairs of unfamiliar wasps, subordinate wasps with experimentally altered facial features ('cheaters') received considerably more aggression from the dominant than did sham controls, indicating that facial patterns are signals and that dishonest signalling imposes social costs.

Variation in the number, size and shape of black spots on the clypeus of *P. dominulus* (Box 1) is remarkable (Fig. 1). Given the signalling potential¹³ of these facial patterns, we tested whether they are related to quality by evaluating their relationship to body size (an important predictor of dominance^{14,15}) and social dominance as determined by pair-wise contests between unfamiliar wasps.

Clypeus spots are strongly correlated with overall body size. Multiple regression yielded both spot number and percentage of clypeus pigmented black as significant predictors of head width (Fig. 2, total regression: $N = 158$, $F = 5.86$, $r^2 = 0.070$, $P = 0.0035$; spot number: $t = 2.07$, $P = 0.040$; percentage of clypeus black: $t = 2.01$, $P = 0.046$). In *Polistes*, head width is the measure of body size most highly correlated with other body size measurements¹⁶.

Facial patterns also predict social dominance beyond their relationship with body size. To control for the effects of body size on dominance, we paired similarly sized unfamiliar foundresses and observed their behaviour as they battled for dominance (similarity in mass between pairs: $r^2 = 0.83$). The alpha (dominant) wasp was easily identified by 'mount' displays, where the beta wasp lowers her antennae and allows the alpha to climb on her head¹⁴. Differences in spot number between contestants significantly predicted dominance, with more dominant wasps having more spots (Table 1). Percentage of clypeus black was not a significant predictor of dominance in the trials, however wasps with relatively less black pigment tended to have higher dominance (Table 1).

We explored the signal value of clypeus patterns in more detail by conducting a pattern analysis on the wasps used in the dominance trials. We quantified the amount and position of black pigment on

each wasp by converting the area of the clypeus containing the population-wide badge variability into a rectangular 30 × 60-pixel bitmap. We then used the bitmap to determine the amount of black pigment present in each of the 60 vertical strips along the horizontal gradient of the clypeus (Fig. 3a). In the overall population, the pigment and pigment variability was concentrated between two prominent peaks of pigment, located on either side of the centre of the clypeus (Fig. 3b). These peaks were much more pronounced in alpha wasps than beta wasps (Fig. 3c), suggesting that more

dominant wasps tend to have more 'broken' facial patterns.

To quantify the difference between alphas and betas, we calculated a 'badge brokenness index' for each wasp as the variability (standard deviation) of amount black pigment deposited along the horizontal gradient located between the two peaks. Overall, mean badge brokenness was significantly higher in alphas than in betas (Fig. 3d, $t_{120} = 3.35$, $P = 0.0011$). Furthermore, badge brokenness was the only significant predictor of dominance when included with spot number and percentage of clypeus black in a multiple logistic regression (as computed in Table 1: whole model, $r^2 = 0.105$, chi-square = 8.86, d.f. = 1, $P = 0.003$; Brokenness, Wald chi-square = 7.52, $P = 0.006$). The high dominance (Table 1) and high brokenness index (Fig. 3d) of ≥ 2 -spot wasps contributed strongly to this relationship, but brokenness also predicted dominance in contests between pairs of 1-spot wasps ($N = 16$ trials, Wald chi-square = 4.79, $P_{(1-tailed)} = 0.044$). This suggests brokenness is a more general predictor of dominance than spot number *per se*. Badge brokenness is also correlated with body size: in the random sample of wasps collected for morphological analysis (see Fig. 2), brokenness was significantly correlated with head width ($N = 158$, $r^2 = 0.028$, $F = 4.53$, $P = 0.035$). The brokenness index is a particularly useful parameter because it collapses a wasp badge's degree of advertised dominance into a singular variable.

Clypeus patterns in *P. dominulus* share four similarities with typical melanin-based badges of status in birds (Box 1). They are: (1) variable, (2) visible¹³ (located on the part of the body most apparent during face-to-face aggressive encounters), (3) associated with body size and (4) associated with dominance. What keeps the relationship between facial patterns and dominance 'honest'? Clypeus spots (most likely eumelanin¹⁷) seem to have low differential production costs because the spots account for less than 1% of the total black pigmentation of *P. dominulus* and the rest of the pigment seems identical from wasp to wasp.

Box 1

Badges of status, social costs and wasps

Badges of status The classic examples are signals used to settle minor dominance contests in flocks of birds⁵. For example, the size of melanin-based throat patches in a variety of species reflects aggression during feeding^{6,24}. However, badges of status are also used in other contexts. Territorial defence is related to the size of the white forehead patch in collared flycatchers *Ficedula albicollis*²⁵ and the redness of the carotenoid-based epaulets in red-shouldered widowbirds *Euplectes axillaris*²⁶. Badges of status appear to be honest (that is, reliable predictors of dominance), but it is unclear how the honesty of these signals is maintained^{6,8}. Indeed they are often argued to have low production costs^{5,8}: for example, the pigment-less badge of the collared flycatcher is probably based on the *cheapest* form of feather development. Without some type of cost, these signalling systems would be vulnerable to the evolutionary spread of a 'cheater' strategy: subordinate individuals which develop, not necessarily with intent, badges associated with higher dominance.

Social costs One hypothesis is that honesty can be maintained if cheaters suffer social costs that outweigh any benefits of cheating^{5-7,9-12}. The main line of support for social costs comes from studies demonstrating that individuals with experimentally altered badges receive more aggression than honest individuals^{6,7,9}. However, many studies have failed to find evidence for increased aggression towards experimentally created cheaters (see ref. 6 for review), and the positive support is rife with problems stemming from difficult to interpret experiments (see refs 6, 27) and potential confounds such as the signalling of individual identity^{4,13,24,28,29}, age-group^{24,29,30} or gender^{24,29,30} rather than status *per se*. In this study we address these problems by evaluating behaviour in both participants of pair-wise dominance trials between unfamiliar contestants of the same age and sex. Social costs are the major hypothesis to explain the honesty of badges of status in birds⁶ and many other taxa⁷, although they have yet to be demonstrated conclusively^{6,7}.

Polistes dominulus Dominance is a key feature in the lives of these common eusocial insects¹⁴. After overwintering, newly emerged queens found new colonies, often co-operatively with other co-foundresses¹⁵. Foundresses fight vigorously to establish dominance rank because dominance determines the amount of reproduction each foundress secures¹⁴. Dominance probably has an important role in settling conflict in many other contexts including the order of queen succession, division of labour, sharing of food and the probability of becoming a future queen. Among insects, these paper wasps are good candidates to evolve visually based badges of status because they are visually acute, diurnal, open living and highly social.

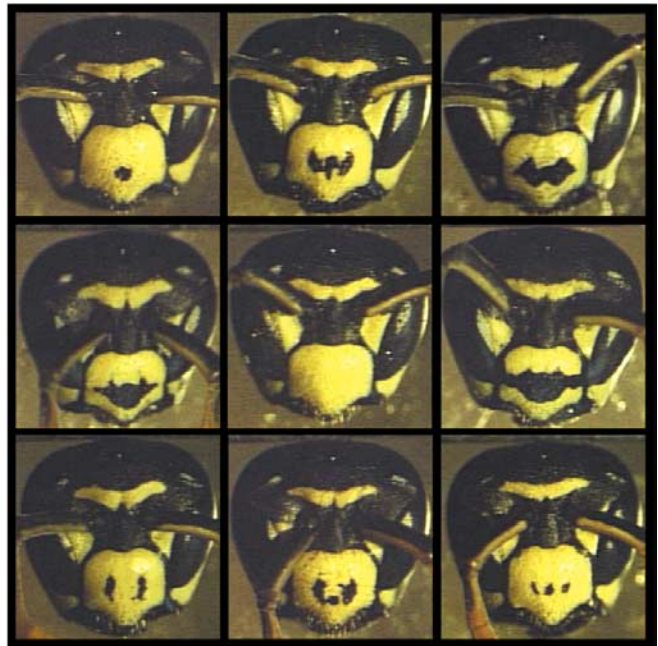


Figure 1 Portraits of nine *P. dominulus* foundresses collected in Ithaca, New York, representing some of the diversity in facial patterns. The central wasp has no black clypeus pigmentation, the remaining top 5 wasps have 1 black spot each, and the bottom 3 wasps have 2, 2 and 3 spots, respectively. In 158 randomly collected foundresses, 19.6% of foundresses had an entirely yellow clypeus, 65.8% had a single black spot, 12.7% had two spots and 1.9% had three spots. On average 13% (± 11 s.d.) of a wasp's clypeus was pigmented black, and this value ranged broadly from 0 to 39%.

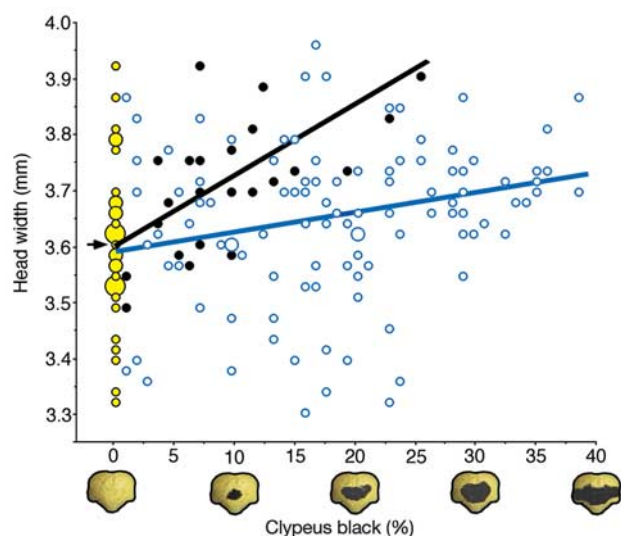


Figure 2 Head width versus clypeus variability. 158 foundresses were collected from 40 single-foundress and 49 multiple-foundress colonies across Tompkins County, New York (May 2001). Model 1 regression in wasps with ≥ 2 -spots, $y = 1.096x + 3.605$, $N = 23$, $r^2 = 0.36$, $F = 12.14$, $P = 0.002$ (black line and black filled circles). Wasps with 1 spot, $y = 0.302x + 3.590$, $N = 104$, $r^2 = 0.046$, $F = 4.94$, $P = 0.029$ (blue line and blue open circles). Mean ($= 3.604$ mm) (arrow) and distribution (yellow circles), in wasps with no spots. The slope of the ≥ 2 -spot regression line is $3.6 \times$ steeper than the 1-spot line, however this difference is not statistically significant (Student's t -test for slopes, $t = 1.75$, $P = 0.08$). Increasing point sizes correspond with one, two and three overlapping data points, respectively.

We tested whether black spots had social costs (that is, costs associated with signal maintenance and accrued during repeated agonistic interactions (Box 1)). We evaluated two critical tests of the hypothesis. Firstly, do alphas react to natural 'cheaters', that is, subordinate wasps who possess badges associated with higher dominance? We specifically predicted that subordinates with higher brokenness indices should receive more aggression from dominants than subordinates with lower brokenness indices. Secondly, do wasps react to experimentally created 'cheaters'^{9,13}? We specifically predicted that wasps whose badges are altered so they no longer reflect their bearers' true behavioural dominance should receive more aggression from opponents than wasps whose badge appearances were not changed.

To test responses to natural cheaters, we scored aggression between contestants after dominance was established (after one wasp submitted to a mount attempt by the other). Alphas reinforced their dominance through repeated mounts of betas. As predicted, the post-dominance mount rate of alphas was strongly and positively correlated with the badge brokenness of betas (Fig. 4). Therefore, after dominance has been determined, subordinate wasps pay higher social costs if they have badges associated with higher dominance.

To test responses to experimentally created cheaters, we conducted a second set of dominance trials (as above). In these trials, the badge of one wasp from each pair was experimentally manipulated¹³ into one of three treatment groups: (1) sham controls were

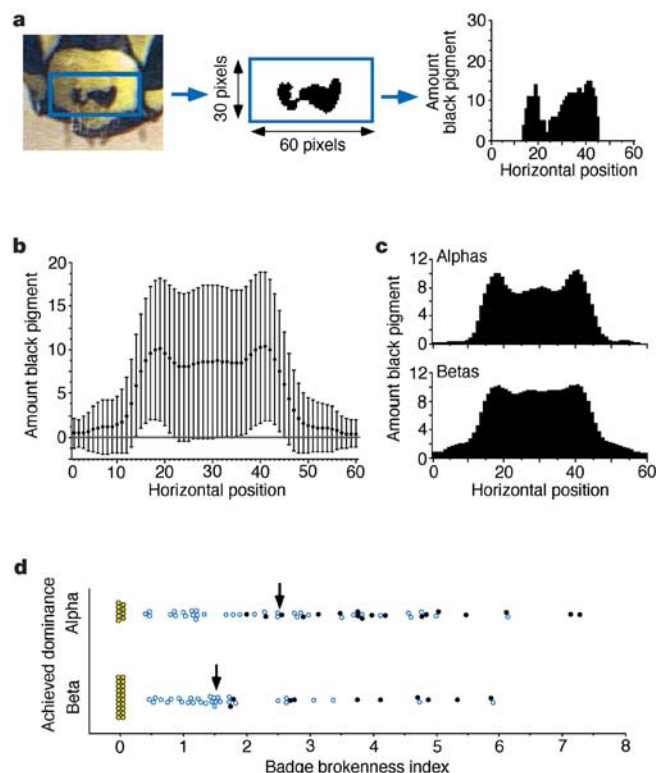


Figure 3 Relationship between pigment deposition and dominance in unmanipulated wasps (paired by mass). **a**, Example of pattern analysis, the badge of each wasp was converted into a 30×60 pixel bitmap to quantify the position and amount of black pigment. **b**, Mean amount of pigment ($\pm$ s.d.) along the horizontal gradient of the badge bitmaps for wasps used in the dominance trials ($N = 122$). **c**, Mean pigment amounts for alphas ($N = 61$) and betas ($N = 61$). **d**, Achieved dominance as a function of badge 'brokenness', a measure of variability in black pigment along the central third of the clypeus ($N = 122$). Arrows denote means. The colour of each point corresponds with spot number: yellow = no spot, blue = 1-spot and black = ≥ 2 -spots.

painted on the clypeus without altering their visual appearance, (2) negative and (3) positive 'cheaters' were painted on the clypeus to reduce, or increase, respectively, their apparent badge rank. The other wasp in each pair was not manipulated (hereafter, 'unpainted').

The manipulation had little impact on behaviour before dominance establishment. Painting did not influence achieved dominance rank (in 29 of 72 trials the manipulated wasp achieved alpha position, chi-square = 2.72, d.f. = 1, $P = 0.10$), and there was no effect of treatment group on probability of becoming dominant (chi-square = 0.41, d.f. = 2, $P > 0.50$). Thus, positive cheaters were no more likely to be alpha than controls or negative cheaters. Threat rates (Fig. 5a) and mount-attempt rates (Fig. 5b) were similar in manipulated and unpainted wasps and did not depend on treatment.

Post-dominance behaviour was affected strongly by the badge alterations. Beta cheaters received significantly more post-dominance mounts from alphas than did beta sham controls (Fig. 5c). Betas manipulated to advertise high dominance received approxi-

Table 1 Mean spot number and percentage of clypeus pigmented black in winners versus losers of dominance contestants (paired by mass).

	Mean ($\pm$ s.e.) of winners	Mean ($\pm$ s.e.) of losers	Wald chi-square*	P
Spot number	1.230 $\pm$ 0.013	0.853 $\pm$ 0.012	7.90	0.010
Percentage of clypeus black	10.4 $\pm$ 0.2	13.4 $\pm$ 0.2	3.28	0.070

*logistical multiple regression of the signed pair-wise differences of these variables as predictors of dominance. The dependent variable was whether a single focal wasp of each pair (assigned randomly) was winner (1) or loser (0). The interaction term between independents was not significant so we removed it from the model. The final regression was significant: $N = 61$ trials, $r^2 = 0.119$, likelihood ratio test chi-square = 10.07, d.f. = 3, $P = 0.007$. When differences in percentage of clypeus black were also removed from the model in step-wise fashion, differences in spot number remained a significant predictor of dominance (Wald chi-square = 5.59, $P = 0.018$).

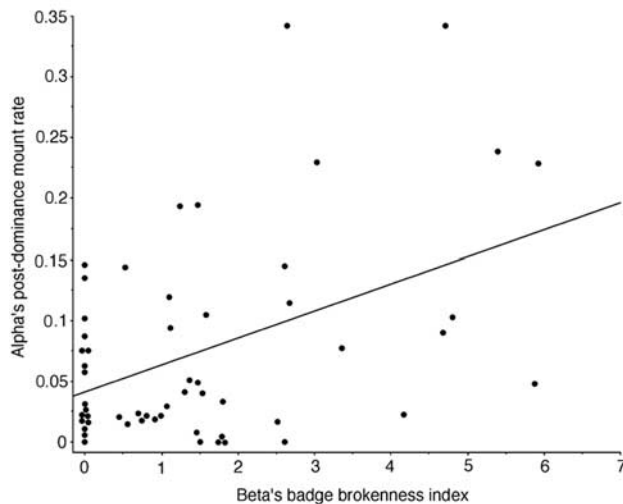


Figure 4 Relationship between post-dominance mount rate by alpha versus beta's badge brokenness. Rates were recorded in unmanipulated wasp contestants (paired by mass) and are reported as $\log(\text{events}/\text{min} + 1)$. Model 1 regression: $y = 0.221x + 0.041$, $N = 55$ trials, $r^2 = 0.18$, $F = 12.97$, $P = 0.0007$.

mately six times more aggression than beta sham controls, and roughly twice the aggression of wasps painted to advertise low dominance. In strong contrast, manipulated wasps who became alpha were similarly aggressive to the unpainted beta independent of their treatment group (Fig. 5d). Furthermore, their aggression levels were similar to those expressed by unpainted alphas towards beta sham controls (analysis of variance (ANOVA), $F_{3,42} = 0.97$, $P = 0.42$). Thus badge alteration did not influence how manipulated individuals treated their opponents, but it did influence strongly how they were treated by these opponents. Interestingly, unpainted betas challenged cheating alphas more frequently than sham control alphas (Fig. 5c). Finally, dominance 'switches' (where the beta successfully claimed alpha position after initial dominance was established) were only observed ($N = 10$) in cheater treatments (Fisher exact $P = 0.002$). Thus facial pattern alteration interfered with accurate exchange of dominance information between contestants, indicating that the facial patterns are a signal.

These results suggest that dishonest status advertisement in *P. dominulus* is kept in check through social costs. Wasps whose badges dishonestly signal their bearers' quality receive more aggression. Increased aggression is costly¹⁸ because it imposes physiological and time costs (for example, see ref. 19), and presumably increases the risk of injury, thereby reducing group productivity (for example, see ref. 20). Because our experimental design allows us to exclude confounding hypotheses (see Box 1), these results represent the most convincing evidence for social costs to date. Although subordinates with increased badge quality received the most aggression, subordinates with reduced badge quality also received more aggression than sham control subordinates. In both cases, the increased aggression probably results from the alpha's detection of a greater incongruence⁹ between badge phenotype and other quality cues (for example, behavioural and/or pheromonal²¹) in the cheater subordinates.

The visual signal of status is probably important in many contexts, as dominance is such a critical aspect of the social behaviour of *P. dominulus*¹⁴. Social costs seem strongest after dominance establishment, so fitness benefits associated with honest signals are probably felt then too: through the signals' effects on conflict settlement regarding dominance succession, relative reproduction, sharing of resources and division of labour.

Wasps with multiple spots (or more generally, more broken patterns) are more dominant, suggesting that these phenotypes

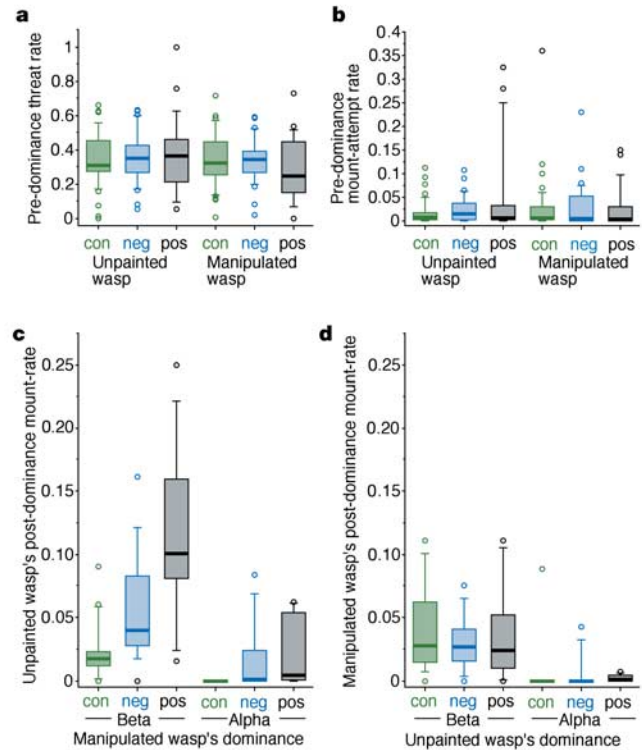


Figure 5 Dominance behaviour in trials where one wasp's facial appearance was experimentally manipulated. Con, sham control; neg, negative cheaters; pos, positive cheaters. Rates are reported and analysed as $\log(\text{events}/\text{min} + 1)$. Box plots show medians, 10th, 25th, 75th and 90th percentiles and outlying data points. **a, b**, Threat rates (**a**) and mount-attempt rates (**b**) were similar in unpainted and manipulated wasps before dominance establishment (ANOVA: threat rates, $F_{5,186} = 0.61$, $P = 0.70$; mount-attempt rates, $F_{5,186} = 1.21$, $P = 0.30$). **c**, Post-dominance mount rates by unpainted wasps towards manipulated wasps. Analysed with two-factor ANOVA using treatment and unpainted wasp's dominance (alpha or beta) as independent variables. Mount rate varied significantly between treatments ($F_{2,65} = 11.12$, $P < 0.0001$). **d**, Post-dominance mount rates by manipulated wasps towards unpainted wasps were similar among treatments ($F_{2,65} = 0.34$, $P = 0.72$).

are more costly. This observation is not consistent with high production costs of pigmentation because wasps with one spot (which tend to be less dominant) have about twice as much black pigment as wasps with ≥ 2 spots (mean percentage of clypeus black in one-spots = 18.4%, in ≥ 2 spots = 9.6%, $t_{125} = 4.09$, $P < 0.0001$). However, this observation is consistent with social costs. An important difference between social costs and production costs is that social costs depend directly on variance in how the signal is perceived (that is, its obviousness) rather than on variance in the absolute state of the signal (for example, its size). Researchers have long known that the hymenopteran compound-eye/visual system is particularly attuned to the amount of disruption in a pattern²². Therefore, the most 'broken' facial patterns may be the most visually apparent phenotypes to receivers. Wasps with more broken patterns may be broadcasting a more obvious signal, and hence have a phenotype associated with higher social costs. □

Methods

Morphology

We mounted the head of each wasp against a ruler affixed to the viewing tray of a $\times 40$ -dissecting microscope. The faces were videotaped with a digital video camera mounted onto the microscope using a Sony CCD camera. We analysed frames from these videos imported into Adobe Photoshop (see ref. 8) to determine (1) the percentage of the clypeus that was pigmented black ($R_1 = 0.99$, $F_{10,11} = 676$, $P < 0.0001$), (2) the number of spots and (3) the maximum width (in mm) of the head capsule ($R_1 = 0.99$, $F_{10,11} = 236$, $P < 0.0001$) (repeatabilities (R_i) calculated using different images of the same wasps).

Size analysis

Because multiple foundresses are often sisters²³, the size analysis (Fig. 2) is not completely independent of genetic relatedness. However, there were no differences between wasps from single and multiple foundress colonies in head width ($t_{156} = -0.66$, $P = 0.51$), spot number ($t_{156} = -0.51$, $P = 0.61$) or percentage of clypeus black ($t_{156} = -1.57$, $P = 0.12$). Furthermore, when mean colony values were used in multiple regression, spot number (but not percentage of clypeus black) was still a significant predictor of head width (total regression: $N = 89$ colonies, $r^2 = 0.09$, $F = 4.29$, $P = 0.017$; spot number: $t = 2.08$, $P = 0.040$; percentage of clypeus black: $t = 1.16$, $P = 0.25$).

Dominance trials

144 recently emerged foundresses were collected in Ithaca, New York (April–May 2002), weighed, marked for individual identification with enamel paint on the thorax and isolated in a small holding container for 0.5–6 h until their trial began. For each trial ($N = 72$), two wasps collected from sites >5 km apart were matched for weight, placed in a $7.5 \times 12.5 \times 4$ cm plastic arena, and their interactions were videotaped for 2 h. After the trial, the contestants' faces were photographed and scored with Photoshop. Videos were scored for two parameters: (1) Dominance: foundresses battled for dominance using the natural range of dominance interactions. *Polistes* wasps aggressively oppose attempted mounts by subordinate wasps, but lower their antennae and remain still to allow mounting by dominants¹⁴. Dominance was considered established when one wasp (= beta) lowered her antennae and allowed the other wasp (= alpha) to mount her. 11 trials where the apparent subordinate never submitted were excluded from further analyses. (2) Post-dominance aggression: to test our first prediction, we recorded the number of mounts subordinates received after dominance was established. Because the interval of post-dominance behaviour varied between trials depending on when dominance was established ($N = 55$, $x = 59.6$ min $\pm$ 31.6 s.d., range = 5.0 to 121.5), our measure of post-dominance mount rates could be affected if mount rates change with time. However mount rate was not related to post-dominance observation time ($r^2 = 0.00$, $F = 0.029$, $P = 0.87$). Six trials where alpha's identity was determined, but we did not obtain videos, were excluded from the analyses of post-dominance behaviour.

Manipulation experiments

192 recently emerged foundresses were collected in Ithaca, New York (April–May, 2003). Pairs of wasps (collected from sites >5 km apart and similar in mass (r^2 between pairs = 0.92)) comprised an unpainted (unmanipulated), and a painted (manipulated)¹³ individual. Each unpainted wasp had a single clypeus spot (indicating average dominance) and received a silver dot on her back for identification. Painted wasps had a range of facial patterns, from 0 to 3 spots, and were placed into one of three treatment groups: (1) 'sham controls' ($N = 44$) were painted without altering their facial appearance using yellow and/or black Testors enamel paint (no.1114 and no.1147, respectively); (2) 'positive cheaters' ($N = 23$) had their original 0 or 1 spot markings altered with black and/or yellow paint to 2 spots; and (3) 'negative cheaters' ($N = 29$) had their original 1 or 2 spot markings altered with yellow paint to 0 spots. The paint was allowed to dry for 20 min before the two wasps were placed in an arena and videotaped for 2 h (as above). Tapes were later analysed for dominance behaviour by E.A.T., blind to treatment group. In each trial, we recorded: (1) dominance (as above; 24 trials where there was no submission by the apparent subordinate were excluded from analyses requiring alpha's identity); (2) pre-dominance threat rate (threat = the initiator lunges towards the recipient) and mount-attempt rate by each wasp; and (3) post-dominance mount rate by each wasp. Post-dominance observation time ($N = 71$, $x = 92.3$ min $\pm$ 27.2 s.d., range = 17.9 to 120.8) was not related to mount rate in all treatments pooled ($r^2 = 0.00$, $F = 0.001$, $P = 0.97$) or in treatments analysed separately (all $P > 0.17$). One trial where there was only one minute of post-dominance behaviour was excluded from the mount rate analysis. In 10 trials (5 in each cheater treatment), the beta successfully assumed the dominant position as indicated by successfully mounting the alpha (that is, alpha submitted).

Received 8 June; accepted 18 August 2004; doi:10.1038/nature02949.

- Zahavi, A. Mate selection—a selection for handicap. *J. Theor. Biol.* **53**, 205–214 (1975).
- Grafen, A. Biological signals as handicaps. *J. Theor. Biol.* **144**, 517–546 (1990).
- Andersson, M. *Sexual Selection* (Princeton Univ. Press, Princeton, New Jersey, 1994).
- Dale, J., Lank, D. B. & Reeve, H. K. Signaling individual identity versus quality: a model and case studies with ruffs, queleas and house finches. *Am. Nat.* **158**, 75–86 (2001).
- Rohwer, S. The social significance of avian winter plumage variability. *Evolution* **29**, 593–610 (1975).
- Senar, J. C. Plumage colouration as a signal of social status. *Proc. Int. Ornithol. Congr.* **22**, 1669–1686 (1999).
- Whiting, M. J., Nagy, K. A. & Bateman, P. W. in *Lizard Social Behavior* (eds Fox, S. F., McCoy, J. K. & Bairds, T. A.) 47–82 (John Hopkins Univ. Press, Baltimore, Maryland, 2003).
- McGraw, K. J., Dale, J. & Mackillop, E. A. Social environment during molt and the expression of melanin-based plumage pigmentation in male house sparrows (*Passer domesticus*). *Behav. Ecol. Sociobiol.* **53**, 116–122 (2002).
- Rohwer, S. & Rohwer, F. C. Status signalling in Harris sparrows: experimental deceptions achieved. *Anim. Behav.* **26**, 1012–1022 (1978).
- Maynard Smith, J. & Harper, D. G. C. The evolution of aggression: can selection generate variability? *Phil. Trans. R. Soc. Lond. B* **319**, 557–570 (1988).
- Owens, I. P. F. & Hartley, I. R. Trojan sparrows—evolutionary consequences of dishonest invasion for the badges of status model. *Am. Nat.* **138**, 1187–1205 (1991).
- Johnstone, R. A. & Norris, K. Badges of status and the cost of aggression. *Behav. Ecol. Sociobiol.* **32**, 127–134 (1993).
- Tibbetts, E. A. Visual signals of individual identity in the paper wasp *Polistes fuscatus*. *Proc. R. Soc. Lond. B* **269**, 1423–1428 (2002).
- Röseler, P. F. in *The Social Biology of Wasps* (eds Ross, K. G. & Matthews, R. W.) 309–335 (Cornell Univ. Press, Ithaca, New York, 1991).

- Reeve, H. K. in *The Social Biology of Wasps* (eds Ross, K. G. & Matthews, R. W.) 99–148 (Cornell Univ. Press, Ithaca, New York, 1991).
- Eickworth, K. R. Differential variation of males and females in *Polistes exclamans*. *Evolution* **23**, 391–405 (1969).
- Pereboom, J. J. M. & Biesmeijer, J. C. Thermal constraints for stingless bee foragers: the importance of body size and coloration. *Oecologia* **137**, 42–50 (2003).
- Clutton-Brock, T. H. & Parker, G. A. Punishment in animal societies. *Nature* **373**, 209–216 (1995).
- Gobin, B., Heinze, J., Straetz, M. & Roces, F. The energetic cost of reproductive conflicts in the ant *Pachycondyla obscuricornis*. *J. Insect Physiol.* **49**, 747–752 (2003).
- Cole, B. J. The social behavior of *Leptothorax allardycei* Hymenoptera Formicidae time budgets and the evolution of worker reproduction. *Behav. Ecol. Sociobiol.* **18**, 165–174 (1986).
- Sledge, M. F., Boscara, F. & Turillazzi, S. Cuticular hydrocarbons and reproductive status in the social wasp *Polistes dominulus*. *Behav. Ecol. Sociobiol.* **49**, 401–409 (2001).
- Horridge, G. A. & Zhang, S. W. Pattern vision in honeybees (*Apis mellifera*): flower-like patterns with no predominant orientation. *J. Insect Physiol.* **41**, 681–688 (1995).
- Strassmann, J. E. in *Natural History and Evolution of Paper-wasps* (eds Turillazzi, S. & West-Eberhard, M. J.) 190–201 (Oxford Univ. Press, Oxford, 1996).
- Whitfield, D. P. Plumage variability, status signaling and individual recognition in avian flocks. *Trends Ecol. Evol.* **2**, 13–18 (1987).
- Pärt, T. & Qvarnstrom, A. Badge size in collared flycatchers predicts outcome of male competition over territories. *Anim. Behav.* **54**, 893–899 (1997).
- Pryke, S. R. & Andersson, S. Carotenoid-based status signalling in red-shouldered widowbirds (*Euplectes axillaris*): epaulet size and redness affect captive and territorial competition. *Behav. Ecol. Sociobiol.* **53**, 393–401 (2003).
- Holberton, R. L., Able, K. P. & Wingfield, J. C. Status signalling in dark-eyed juncos, *Junco hyemalis*: plumage manipulations and hormonal correlates of dominance. *Anim. Behav.* **37**, 681–689 (1987).
- Shields, W. M. The social significance of avian winter plumage variability: a comment. *Evolution* **31**, 905–907 (1977).
- Watt, D. J. A comparative study of status signalling in sparrows (genus *Zonotrichia*). *Anim. Behav.* **34**, 1–15 (1986).
- Balph, M. H., Balph, D. F. & Romesburg, H. C. Social status signalling in winter flocking birds: an examination of a current hypothesis. *Auk* **96**, 78–93 (1979).

Acknowledgements Thanks to M. Andrade, P. Buston, B. Daley, C. Gilbert, A. Houde, D. Lank, K. McGraw, H. K. Reeve, T. Seeley, P. Sherman, W. Gronenberg and the FAB evolution group (S.F.U.) for helpful comments on the manuscript, D. Lank for assistance with the statistical analysis, W. Gronenberg and H. K. Reeve for research support and L. Marakon for research assistance. E.A.T. was supported by a NSF graduate fellowship and NIH Training Grant. J.D. was supported by a N.S.E.R.C. post-doctoral fellowship.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to E.A.T. (eat11@email.arizona.edu).

Environmental biosafety and transgenic potato in a centre of diversity for this crop

Carolina Celis^{1*}, Maria Scurrah^{2*}, Sue Cowgill^{3*}, Susana Chumbiauca^{2*}, Jayne Green^{3*}, Javier Franco⁴, Gladys Main⁴, Daan Kiezebrink³, Richard G. F. Visser¹ & Howard J. Atkinson³

¹Laboratory of Plant Breeding, Wageningen University, PO Box 386, 6700 AJ, Wageningen, The Netherlands

²SENASA, Pedro de Pasaje Francisco de Zela No. 150 piso 10, Ministerio de Agricultura, Lima 11, Lima, Peru

³Centre for Plant Sciences, University of Leeds, Leeds LS2 9JT, UK

⁴Fundación PROINPA, Av. Blanco Galindo km 121/2, PO Box 4285, Cochabamba, Bolivia

* These authors contributed equally to this work

The Nuffield Council on Bioethics^{1,2} suggests that introgression of genetic material into related species in centres of crop biodiversity is an insufficient justification to bar the use of genetically modified crops in the developing world. They consider that a precautionary approach to forgo the possible benefits invokes the fallacy of thinking that doing nothing is itself without risk to the poor. Here we report findings relevant to this and other aspects of environmental biosafety for genetically modified

potato in its main centre of biodiversity, the central Andes. We studied genetically modified potato clones that provide resistance to nematodes, principal pests of Andean potato crops³. We show that there is no harm to many non-target organisms, but gene flow occurs to wild relatives growing near potato crops. If stable introgression were to result, the fitness of these wild species could be altered. We therefore transformed the male sterile cultivar Revolution to provide a genetically modified nematode-resistant potato to evaluate the benefits that this provides until the possibility of stable introgression to wild relatives is determined. Thus, scientific progress is possible without compromise to the precautionary principle.

A genetically modified potato can be developed within 5 years rather than the 8–15 years often needed for conventional breeding⁴. This rapidity offers hope of countering within the next decade many of the biotic and abiotic stresses that affect subsistence potato growers. Genetically modified nematode-resistant (GMNR) potato clones that express a cysteine proteinase inhibitor (cystatin) are an example of a beneficial trait for subsistence growers³. The expression of cystatin from rice seeds in roots controls potato cyst nematode (*Globodera* spp.) by impairing digestion of its dietary protein^{5,6}. These GMNR potato clones have been examined for several aspects of biosafety that are relevant wherever they are grown. They are not detrimental to above-ground insect associates of the crop^{7,8}, their parasitoid natural enemies⁹ or either soil fauna or soil microbial communities¹⁰. Furthermore, expression of the transgene can be limited to roots¹¹ and there is no food risk to humans¹². However, a

specific issue surrounds the future use of these transgenic potato clones in South and Central America. This is a centre of origin for the crop, with 130 wild potato species (*Solanum* L. section *Petota* Dumort.) recognized in Peru and Bolivia¹³. When a genetically modified crop grows in close proximity to sexually compatible wild relatives, the biosafety of transgenes must be considered on a case-by-case basis¹⁴.

We undertook a study of the non-target effects of GMNR potato before seeking field trials for such clones in South America. We determined the community-level physiological profile of rhizosphere bacteria at harvest by using methods that have already detected such effects for transgenic potato expressing a lectin¹⁵. Grown in pots in a containment glasshouse, GMNR potato cultivars developed for future use in Bolivia imposed no change on this profile relative to their untransformed parents (Fig. 1a). This lack of effect of cystatin contrasts with our results for rhizosphere soil recovered from different conventional crops growing in an experimental field plot in Bolivia (Fig. 1b–d). The soil associated with native cultivars grouped together but differed from that in which potato cv. Desiree or either broad bean or lupin were growing. Amino acids, carbohydrates and phenolic substrates contributed most to this discrimination (Supplementary Information). Such an effect of plant species has been reported previously¹⁶.

We also established in a UK field trial that GMNR potato do not influence the density or number of soil nematode species beneath a mature potato crop. The GMNR potato did not disturb the maturity index¹⁷ on the basis of the proportion of colonizer and persister species in the soil nematode community (Supplementary Information). This contrasted with the effect of applying pre-plant nematicide that subsequently favoured fast growing species several months later (Fig. 1e). This and our previous work^{7–10} suggest that soil organism populations will not be harmed at a trial site. We also propose that the needs of the precautionary principle would be met for any subsequent wider uptake if genetically modified plants impose no more impact on the soil environment than do common agricultural practices.

To establish whether gene flow would occur from GMNR field trials, we selected six wild species on the basis of a preliminary

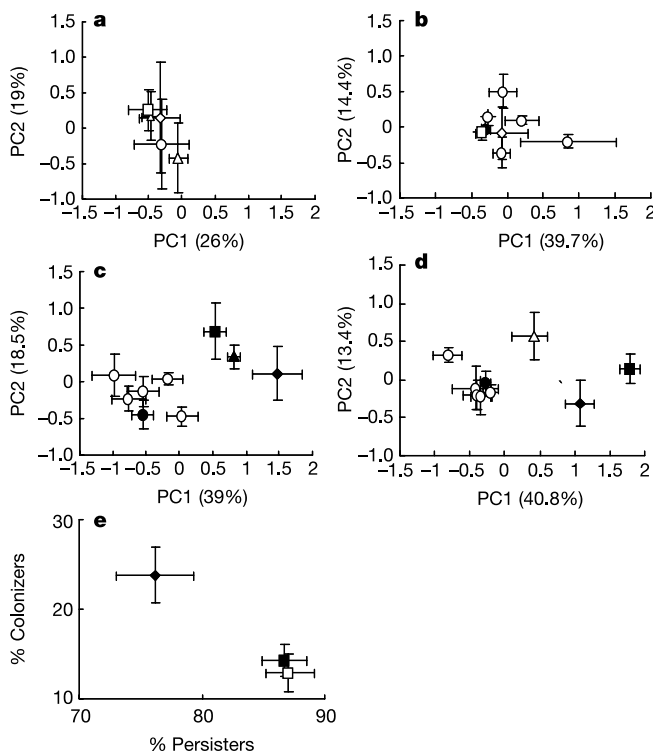


Figure 1 Studies on rhizosphere organisms. Shown are principal component (PC 1 and PC 2) scores¹⁵ summarizing community-level physiological profiles for microorganisms. **a**, Untransformed and transgenic containment grown cv. Maria Huanca (open squares and circles) and cv. Revolution (open diamonds and triangles). **b–d**, Legumes and non-transgenic cv. Revolution (filled circles), cv. Desiree (diamonds) and other potatoes (open circles) in a Bolivian field at hilling, flowering and harvesting. **c**, Revolution differed in PC 1 from filled versions of symbols for cv. Desiree and broad bean (square) in **c** and **d**, and from lupin (triangle) in **c**. **e**, Similar persister and colonizer proportions for nematode species associated with untransformed (filled square) and transgenic (open square) cv. Desiree are altered by a pre-plant nematicide (filled diamond). $P < 0.05$ in **c–e** by one-way analysis of variance, Tamhane test. Values are the mean $\pm$ s.e.m.

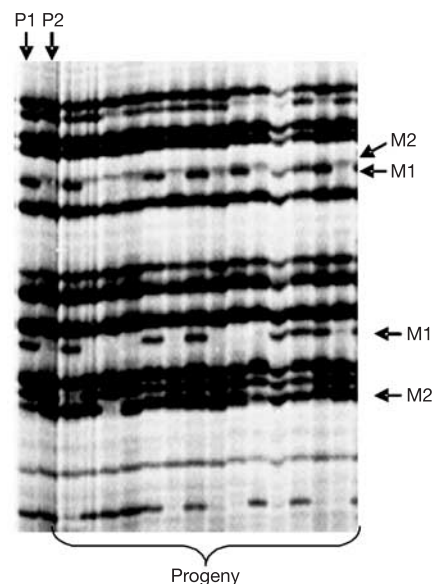


Figure 2 Confirmation of hybrid progeny resulting from forced crosses. Shown is an example autoradiograph confirming, by molecular characterization, the formation of hybrid progeny resulting from forced crosses, in this case between the wild species *S. raphanifolium* (P1, female parent) and *S. tuberosum* Andigenum cv. Qompis (P2, pollen donor). M1 and M2 indicate example AFLP bands specific to P1 and P2, respectively. All progeny shown are hybrids.

Table 1 Number of hybrid seedlings per total seedlings identified by AFLP fingerprinting

Expt*	<i>S. tuberosum</i> 4 × (4EBN)	Pollination	<i>S. albicans</i> 6 × (4EBN)	<i>S. acaule</i> 4 × (2EBN)	<i>S. chomathophilum</i> 2 × (2EBN)	<i>S. raphanifolium</i> 2 × (2EBN)	<i>S. bukasovii</i> 2 × (2EBN)	<i>S. sparsipilum</i> 2 × (2EBN)
A	Male	Forced	0/317	0/169	ND	48/48	ND	168/168
A	Female	Forced	405/548	93/161	3/291	3/90	286/640	30/122
B	Male	Open	ND	15/221	ND	ND	9/340	17/569
Seeds/plant			ND	352 ± 110	ND	ND	257 ± 156	528 ± 237
% viable			ND	80 ± 16%	ND	ND	68 ± 10%	61 ± 12%

*In experiment A, the number of hybrid seedlings per total seedlings examined was determined by AFLP analysis from a seed set obtained after reciprocal hand crossing of wild *Solanum* species with either *S. tuberosum* Andigenum cultivars or hybrids of these and Tuberousum cultivars. In experiment B, the same analysis was used for a seed set obtained from wild *Solanum* spp. in an open pollinated field experiment with potato plants to identify progeny for which the pollen donor was *S. tuberosum*. The seeds produced per plant in experiment B and their percentage viability are given for three species that produced abundant seed. ND, not determined owing to lack of seeds. Measurements of error are mean ± s.e.m.

survey of four agro-ecological zones in Peru to define examples of wild *Solanum* species that occur close to potato plots. First, we confirmed that our accessions did cross, as reported before in some cases¹⁸, by using forced, hand pollinations between them and *Solanum tuberosum* cultivar group Andigenum landraces¹⁹ and improved cultivars. The seedlings from the set seed were analysed using amplified fragment length polymorphism (AFLP) markers and hybrids involving all six wild species were identified (Fig. 2).

As hand pollination may not represent natural cross-pollination events reliably, we established five open pollination field trials in different agro-ecological zones at Puno, Junín, Cajamarca and Cusco. AFLP established that the progeny from each of three wild relatives with accessions in the trial included a few seedlings that were fathered by *S. tuberosum* cultivar group Andigenum (Table 1). This confirms that gene flow can occur from a cultivated potato to wild relatives in the field. Assuming that the nearest plant of the male parent identified by AFLP fingerprinting was the pollen donor, their mean distance was less than three plants from the pollen recipient. This is consistent with previous work that established that 2% of seedlings were transgenic when the cross-pollination distance from non-transgenic females to transgenic pollen donors was 3 m and 24% were transgenic when the parents grew next to each other²⁰.

Concurrent flowering of the potato plants and *Solanum acaule*, *Solanum bukasovii* and *Solanum sparsipilum* spanned 52 d. Even the value of 32 d for the early flowering *Solanum megistacrobolobum* provided ample opportunity for cross-pollination by insects. Bumblebees and their relatives are the main pollinators of *Solanum*²¹. Pollen is the only food reward offered by *Solanum* spp. and they are not frequently visited by honey bees seeking nectar. In addition, the anthers of these plants require sonication by insects to release pollen, and thus the range of pollinating insects is restricted. Bumblebees typically forage over 70–631 m (ref. 22), but pollen

from one flower is usually deposited only across a limited number that are subsequently visited. This and factors such as residence time in one crop favours highly localized cross-pollination of plants near the pollen source²³.

We studied visits by insects to flowers at five sites in Peru. Twelve bee species were recorded, but species at the sites differed and only the non-abundant *Bombus funebris* was present at all sites. Results at Puno provide an example. There, the black Andean bee (*Lonchopria* spp.) and *Bombus opifex* and *B. funebris* provided, respectively, 36.9 ± 5.4%, 34.7 ± 4.8% and 4.4 ± 0.2% (mean ± s.e.m.) of the total flower visits. The abundance of bees also varied with sites in counts made for 10 min of each daylight hour at least once a week throughout the field trial. At the Puno site, the grand mean was 505 ± 12 bees per day, whereas it was only 107 ± 27.2 bees per day at the site in Junín. The extent of pollen flow between *Solanum* species will vary considerably with locality in the central Andes, bumblebee densities and species, their foraging behaviour and the location of wild relatives within and near future transgenic potato crops.

Introgression of a transgene from GMNR potato might confer resistance to nematodes on wild relatives that are currently susceptible to *Globodera* spp. We found that this nematode did reproduce on five of six wild *Solanum* species studied with an increase in number ranging from 4.38 ± 1.40-fold for *S. acaule* to 2.87 ± 1.16-fold (mean ± s.e.m.) for *Solanum raphanifolium*. The exception was *S. sparsipilum* (0.89 ± 0.31-fold for four accessions). Susceptible native and improved cultivars are larger with more substantial root systems than their wild relatives, and they supported higher multiplication of *Globodera* (19.8 ± 2.07-fold; mean value for 24 cultivars). The possibility that a stably introgressed transgene for pest resistance might benefit a wild relative growing near a crop species¹⁴ requires examination on a case-by-case basis. Plant invasiveness²⁴ for both wild and cultivated species of *Solanum* also needs to be considered. The precautionary principle requires such a possibility to be prevented, at least until this concern is evaluated.

The Andean cultivar Revolucion is of particular interest to meet this interim need. It is one of the male sterile cultivars arising from crosses between the *S. tuberosum* Andigenum and Tuberousum cultivar groups. It lacks viable pollen to fertilize eggs of another compatible plant during sexual reproduction. It did not flower abundantly and flower drop occurred at our trial sites. We therefore transformed and developed six GMNR–cystatin lines that provided partial resistance to *Globodera pallida* relative to ex-tissue-culture, untransformed plants of this cultivar (Fig. 3). This level of resistance is similar to that reported before⁵ and is sufficient to prevent yield loss owing to the nematode within the short-rotation courses that necessarily prevail among subsistence growers in Bolivia³. We propose that transgenic cv. Revolucion provides a basis for initial field trials of nematode resistance or other traits of value without gene flow from the potato on trial.

From the current study and previous work, we conclude that there is no basis for invoking the precautionary principle to bar initial GMNR–cystatin potato trials for their impact on non-target soil microbes and fauna or associates of the crop's foliage. No basis for harm has been detected and any effect would be local to the site.

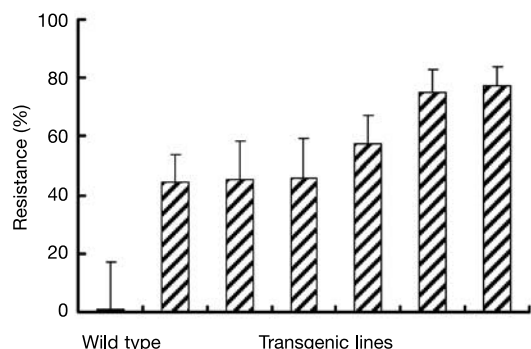


Figure 3 Resistance to *G. pallida* in male-sterile *S. tuberosum* cv. Revolucion. Resistance to *G. pallida* is conferred on susceptible, male-sterile *S. tuberosum* cv. Revolucion when it is transformed to express cystatin by using the construct CaMV35S/OclΔD86 (ref. 5). Resistance is expressed for six transgenic lines in terms of the *G. pallida* multiplication rate that they supported as a percentage of the rate of 15.3 ± 2.5-fold obtained for *G. pallida* multiplication on untransformed cv. Revolucion. Values are mean ± s.e.m.

By contrast, any gene flow from potato to wild *Solanum* species in the Andes that grow nearby may not remain localized and this requires further study. In the mean time, a barrier to outcrossing can be provided by cv. Revolucion. It provides a basis to underpin the development of biosafe genetically modified field trial regulations for potato in the Andes. Transgenic planting should be limited to male sterile cultivars while concerns over possible introgression of any given trait are evaluated over several generations of random mating among individuals of a wild species by defined methods²⁴. When such information is available, the biosafety of a transgenic, male fertile cultivar for that trait can be evaluated. Assessing both benefits and concerns provides an informed scientific basis on which to determine whether the needs of the poor conflict with the precautionary principle for potato in its centres of diversity. □

Methods

Field trial for the effects of GMNR potato on soil nematodes

Tubers were planted on 8 June 2002 at the University of Leeds Farm, UK (DTER consent 98/R31/01) according to a described layout⁵. Soil samples were taken from beneath the plants on 2 August 2002 and bulked for each of the three treatments of GMNR–cystatin plus non-transformed Desiree in soil either not treated or treated with the nematicide aldicarb (Temik at 33.6 kg ha⁻¹; 10% active ingredient) before planting. Nematodes were extracted from the samples by the tray method²⁵, fixed in 2 × TAF²⁵, and counted before 50 individuals per sample were identified to genus. The maturity index was determined as described¹⁷ to separate colonizer from persister species.

Community-level profile of rhizosphere microorganisms

Previously developed GMNR⁶ and control cultivars were pot grown in a containment glasshouse in the UK to provide soil samples. Soils were also sampled from plants in the field by coring into the root zone of five plants per genotype growing in a small field trial at Toralapa Experimental Station, Cochabamba Province, Bolivia. The extraction of soil microorganisms from soil samples and the use of BIOLOG Eco plates were done by standard procedures²⁶. We applied principal component analysis to the data set²⁷.

Hand crossing of *Solanum* species

The various *Solanum* spp. were grown from seed or tubers in glasshouses routinely used by the potato breeders at the International Potato Centre, Peru. Experienced workers carried out hand crossing.

Open-pollination among *Solanum* species

Five open pollination nurseries were planted in Peru at one site in Cajamarca, Cusco and Puno and two sites in Junin. Plants were sown randomly 0.35 by 1.0 m apart with ten replicate plants using several varieties from up to four out of six cultivated *Solanum* species commonly found in these sites. Flower number was counted twice per week. Mature berries were hand-picked and seed was recovered from them in the laboratory.

Field entomology

Once a week in February 2003 and twice a week in March 2003, we counted the flying insects round the perimeter of each plot for 10 min every hour from 0700 to 1800. Voucher specimens of each species were collected and subsequently identified to genus or species. Four varieties were planted in four rows of 6 m in two randomized replicates for insect observation.

AFLP fingerprinting

At least 100 seeds per progeny of both hand and field cross-pollinations were germinated in a containment glasshouse at the Laboratory of Plant Breeding, University of Wageningen, The Netherlands. DNA was isolated from 2–3-week-old, individual seedlings by the CTAB method²⁸ with modifications. The AFLP procedure was done by a standard method²⁹ with the primer combination E35M48, which generated the highest number of polymorphic markers (203) of the 20 primer combinations most commonly used in potato (Potato and Tomato Genetic and Genomic Database, <http://potatodb.dpw.wau.nl/>).

Multiplication of *Globodera* on *Solanum* species

Cysts of *Globodera* were collected from field populations at Puno (Tawaco), Huancayo (Pampa Cruz) and La Libertad (Paraiso). Five replicate tubers per *Solanum* entry were grown individually in 0.2-l clay pots using a sand and peat mixture plus ten viable eggs per gram of soil as cysts in muslin bags. After host senescence, new cysts were extracted by using a Fenwick can and the eggs inside were counted²⁵. We corrected the multiplication rates of *Globodera* for eggs that did not emerge from the parental cysts.

GMNR–cystatin cv. Revolucion

A cystatin construct (CaMV35S/OcIΔD86) was used for *Agrobacterium*-mediated transformation of cv. Revolucion as described for cv. Desiree. We used western blots to select six positive lines expressing more than 0.2% total soluble protein in roots. Transgene presence was confirmed by polymerase chain reaction (PCR) and expression by PCR with reverse transcription as described⁶. Ten plantlets of each transgenic line plus an untransformed ex-tissue-culture line were grown in a containment glasshouse at

18 ± 3 °C and with a 14-h day length for 4 weeks before transplanting into soil containing ten viable eggs per *G. pallida* (cysts of a Toralapa population, Bolivia). After senescence, the haulms were harvested, and the soil was allowed to dry in pots before the post-harvest *G. pallida* population of each was determined as above. The plants grew equally and no phenotypic changes were observed for the transgenic lines.

Received 28 July; accepted 17 September 2004; doi:10.1038/nature03048.

1. Nuffield Bioethics Committee. *Genetically Modified Crops: the Ethical and Social Issues* (Nuffield Bioethics Committee, London, 1999).
2. Nuffield Bioethics Committee. *The Use of Genetically Modified Crops in Developing Countries* (Nuffield Bioethics Committee, London, 2004).
3. Atkinson, H. J., Green, J., Cowgill, S. E. & Levesley, A. The case for genetically modified crops with a poverty focus. *Trends Biotechnol.* **19**, 91–96 (2001).
4. Heeres, P., Schippers-Rozenboom, M., Jacobsen, E. & Visser, R. G. F. Transformation of a large number of potato varieties: genotype-dependent variation in efficiency and somaclonal variability. *Euphytica* **124**, 13–22 (2002).
5. Urwin, P. E., Troth, K. M., Zubko, E. I. & Atkinson, H. J. Effective transgenic resistance to *Globodera pallida* in potato field trials. *Mol. Breed.* **8**, 95–101 (2001).
6. Urwin, P. E., Green, J. & Atkinson, H. J. Expression of a plant cystatin confers partial resistance to *Globodera*, full resistance is achieved by pyramiding a cystatin with natural resistance. *Mol. Breed.* **12**, 263–269 (2003).
7. Cowgill, S. E., Wright, C. & Atkinson, H. J. Transgenic potatoes with enhanced levels of nematode resistance do not have altered susceptibility to non-target aphids. *Mol. Ecol.* **11**, 821–827 (2002).
8. Cowgill, S. E. & Atkinson, H. J. A sequential approach to risk assessment of transgenic plants expressing protease inhibitors: effects on nontarget herbivorous insects. *Transgenic Res.* **12**, 439–449 (2003).
9. Cowgill, S. E., Danks, C. & Atkinson, H. J. Multitrophic interactions involving genetically modified potatoes, nontarget aphids, natural enemies and hyperparasitoids. *Mol. Ecol.* **13**, 639–647 (2004).
10. Cowgill, S. E., Bardgett, R. D., Kiezebrink, D. T. & Atkinson, H. J. The effect of transgenic nematode resistance on non-target organisms in the potato rhizosphere. *J. Appl. Ecol.* **39**, 915–923 (2002).
11. Lilley, C. J., Urwin, P. E., Johnston, K. A. & Atkinson, H. J. Preferential expression of a plant cystatin at nematode feeding sites confers resistance to *Meloidogyne incognita* and *Globodera pallida*. *Plant Biotechnol. J.* **2**, 3–12 (2004).
12. Atkinson, H. J., Johnston, K. A. & Robbins, M. Prima facie evidence that a phytocystatin for transgenic plant resistance to nematodes is not a toxic risk in the human diet. *J. Nutr.* **134**, 431–434 (2004).
13. Spooner, D. M. & Hijmans, R. J. Potato systematics and germplasm collecting 1989–2000. *Am. J. Pot. Res.* **78**, 237–268 (2001).
14. Stewart, C. N., Halfhill, M. D. & Warwick, S. I. Transgene introgression from genetically modified crops to their wild relatives. *Nature Rev. Genet.* **4**, 806–817 (2003).
15. Griffiths, B. S., Goeghegan, I. E. & Robertson, W. M. Testing genetically engineered potato, producing lectins GNA and Con A, on non-target soil organisms and processes. *J. Appl. Ecol.* **37**, 159–170 (2000).
16. Söderberg, K. H., Probanza, A., Jumpponen, A. & Bååth, E. The microbial community in the rhizosphere determined by community-level physiological profiles (CLPP) and direct soil- and cfu-PLFA techniques. *Appl. Soil Ecol.* **25**, 135–145 (2004).
17. Bongers, T. The Maturity Index, the evolution of nematode life history traits, adaptive radiation and cp-scaling. *Plant Soil* **212**, 13–22 (1999).
18. Jackson, S. A. & Hanneman, R. E. Jr Crossability between cultivated and wild tuber- and non-tuber-bearing Solanums. *Euphytica* **109**, 51–67 (1999).
19. Huamán, Z. & Spooner, D. M. Reclassification of landrace populations of cultivated potatoes (*Solanum* sect. *Petota*). *Am. J. Bot.* **89**, 947–965 (2002).
20. McPartlan, H. C. & Dale, P. J. An assessment of gene-transfer by pollen from field-grown transgenic potatoes to nontransgenic potatoes and related species. *Transgenic Res.* **3**, 216–225 (1994).
21. Batra, S. W. T. Male-fertile potato flowers are selectively buzz-pollinated only by *Bombus terrestris* in upstate New York. *J. Kans. Entomol. Soc.* **66**, 252–254 (1993).
22. Osborne, S. J. et al. A landscape-scale study of bumble bee foraging range and constancy, using harmonic radar. *J. Appl. Ecol.* **36**, 519–533 (1999).
23. Cresswell, J. E., Osborne, J. L. & Bell, S. A. A model of pollinator-mediated gene flow between plant populations with numerical solutions for bumblebees pollinating oilseed rape. *Oikos* **98**, 375–384 (2002).
24. Hancock, J. K. A framework for assessing the risk of transgenic crops. *BioScience* **53**, 512–519 (2003).
25. Southey, J. F. *Laboratory Methods for Work with Plant and Soil Nematodes* (Her Majesty's Stationery Office, London, 1986).
26. Garland, J. L., Mills, A. L. & Young, J. S. Relative effectiveness of kinetic analysis vs single point readings for classifying environmental samples based on community-level physiological profiles (CLPP). *Soil Biol. Biochem.* **33**, 1059–1066 (2001).
27. Balser, T. C., Kirchner, J. W. & Firestone, M. K. Methodological variability in microbial community level physiological profiles. *Soil Sci. Soc. Am. J.* **66**, 519–523 (2002).
28. Bernatzky, R. & Tanksley, S. D. Methods for detection of single or low copy sequences in tomato on Southern blots. *Plant Mol. Biol. Rep.* **4**, 37–41 (1986).
29. Vos, P., Hogers, R. & Bleeker, M. AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Res.* **23**, 4407–4414 (1995).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank A. Salas, J. Arcos and R. Canto for assistance with crossing and field experiments; and C. Rasmussen for insect identification. We are grateful to the International Potato Centre and PROINPA for access to their facilities in Peru and Bolivia, respectively. The project was funded by a grant from the European Union.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to H.J.A. (h.j.atkinson@leeds.ac.uk).

A pancreatic islet-specific microRNA regulates insulin secretion

Matthew N. Poy¹, Lena Eliasson³, Jan Krutzfeldt¹, Satoru Kuwajima¹, Xiaosong Ma³, Patrick E. MacDonald³, Sébastien Pfeffer², Thomas Tuschl², Nikolaus Rajewsky⁴, Patrik Rorsman^{3,5} & Markus Stoffel¹

¹Laboratory of Metabolic Diseases and ²Laboratory of RNA Molecular Biology, The Rockefeller University, 1230 York Avenue, New York, New York 10021, USA

³Department of Physiological Sciences, Lund University, SE-221 84 Lund, Sweden

⁴Department of Biology, Biology & Mathematics, New York University, New York, New York 10003, USA

⁵Oxford Centre for Diabetes, Endocrinology and Metabolism, University of Oxford, Churchill Hospital, Oxford OX3 7LJ, UK

MicroRNAs (miRNAs) constitute a growing class of non-coding RNAs that are thought to regulate gene expression by translational repression¹. Several miRNAs in animals exhibit tissue-specific or developmental-stage-specific expression, indicating that they could play important roles in many biological processes^{2–4}. To study the role of miRNAs in pancreatic endocrine cells we cloned and identified a novel, evolutionarily conserved and islet-specific miRNA (*miR-375*). Here we show that overexpression of *miR-375* suppressed glucose-induced insulin secretion, and conversely, inhibition of endogenous *miR-375* function enhanced insulin secretion. The mechanism by which secretion is modified by *miR-375* is independent of changes in glucose metabolism or intracellular Ca²⁺-signalling but correlated with a direct effect on insulin exocytosis. *Myotrophin* (*Mtpn*) was predicted to be and validated as a target of *miR-375*. Inhibition of *Mtpn* by small interfering (si)RNA mimicked the effects of *miR-375* on glucose-stimulated insulin secretion and exocytosis. Thus, *miR-375* is a regulator of insulin secretion and may thereby constitute a novel pharmacological target for the treatment of diabetes.

MicroRNAs (miRNAs) are 21- to 23-nucleotide (nt) non-coding RNAs processed from double-stranded hairpin precursors and have been identified in the genomes of a wide range of multicellular life forms, including plants and animals^{1,5}. The function of miRNAs in vertebrates and mammals is largely unknown, but studies in *Caenorhabditis elegans* and *Drosophila melanogaster* have revealed that miRNAs can bind to target sites in messenger RNAs with imperfect base pairing and, by unknown mechanisms, significantly reduce translational efficiency^{6,7}. Furthermore, genetic studies in these organisms have identified important functions of specific miRNAs in the coordination of cell proliferation and cell death during development and in fat metabolism^{8,9}. To assess the function of miRNAs in regulating metabolism in mammals we have examined endocrine cell types of the pancreas.

We cloned 21- to 23-nt RNAs from total RNA of the glucose-responsive murine pancreatic β -cell line MIN6 and murine pancreatic α -cell line TC1 (ref. 10). We identified 67 different miRNA sequences, 11 of which have not been previously identified and which are conserved in other vertebrates¹¹ (see Supplementary Tables 1 and 2). The microRNA *miR-375* was the most abundant of all novel clones, with an overall abundance of 6.6% and 5.3% in MIN6 and TC1 cells, respectively (see Supplementary Table 1). Northern blot analysis confirmed that expression of *miR-375* was restricted to MIN6 and TC1 cells and mouse pancreatic islets, and not found in other tissues, including exocrine pancreas, liver, lung, fat, intestine, brain, kidney, spleen, heart and testes (Fig. 1a–c, and data not shown). The expression of *miR-376* was limited to MIN6 cells and pancreatic islets. These data suggest that we had identified novel, pancreatic islet-specific miRNAs.

To analyse the function of *miR-375* and *miR-376*, we first

increased the cellular miRNA concentration by introduction of siRNA duplexes homologous in sequence to *miR-375* and *miR-376* (si-375 and si-376, respectively). MIN6 cells were transfected with these siRNAs and the effect on glucose-induced insulin secretion in MIN6 cells was determined. As positive and negative controls, siRNAs targeting the *glucokinase* gene (si-Gck), a key regulator of glucose-stimulated secretion, or *apolipoprotein M* (si-apoM), a gene not expressed in pancreatic β -cells (data not shown), were transfected into MIN6 cells (Fig. 1d, e). Insulin secretion in response to a 25-mM glucose stimulus was decreased in cells transfected with si-Gck and si-375 compared to control si-apoM (Fig. 1d, data not shown). In contrast, an siRNA with mutations in the nucleus of *miR-375* sequence (si-375MUT) had no effect on glucose-stimulated secretion (Fig. 1d). Transfection of synthetic siRNA homologous to several other miRNAs, including *miR-376*, *miR-129*, *miR-130* and *miR-210* had no effect on basal or glucose-stimulated insulin secretion compared to control (data not shown). We used antisense 2'-O-methyl oligoribonucleotides to specifically inhibit miRNAs^{12,13}. A 2'-O-methyl oligoribonucleotide complementary to *miR-375* (2'-O-me-375) was shown to anneal to endogenous *miR-375* in MIN6 cells by competing off detection by a labelled probe (see Supplementary Fig. 1). Transfection of 2'-O-me-375 into MIN6 cells enhanced glucose-stimulated insulin secretion 1.4-fold compared to cells transfected with a control 2'-O-me-eGFP (Fig. 1f). Collectively, these data indicate that *miR-375* is an inhibitor of glucose-stimulated insulin secretion.

To express *miR-375* in primary cells, we generated a recombinant adenovirus expressing *miR-375* (Ad-375). In initial studies, MIN6 cells were infected with a control adenovirus expressing enhanced green fluorescent protein (Ad-eGFP) or increasing concentrations of Ad-375. Northern blot analysis showed a dose-dependent increase of *miR-375* expression (Fig. 2a). Over-expression of *miR-375* in MIN6 cells at a multiplicity of infection (MOI) of 50 led to an ~2.5-fold increase in expression and resulted in an ~40% reduction in insulin secretion induced by 25 mM glucose compared to cells infected with Ad-eGFP (Fig. 2b). The defect in insulin secretion did not result from defective insulin synthesis because total insulin content was equivalent in Ad-375- and Ad-eGFP-infected MIN6 cells (data not shown).

We measured insulin secretion in Ad-375-infected MIN6 cells that were stimulated with 30 mM KCl (to open voltage-gated Ca²⁺ channels; Fig. 2c) or 500 μ M tolbutamide (to close ATP-regulated K⁺ channels and elicit electrical activity; Fig. 2d). Insulin secretion triggered by either of these stimuli was reduced in cells infected with Ad-375 compared to cells infected with Ad-eGFP. Also, total intracellular ATP levels at low or high glucose concentrations (2.8 mM and 25 mM, respectively) were not diminished in Ad-375-infected cells (see Supplementary Fig. 2). Collectively, these data strongly suggest that over-expression of *miR-375* reduces insulin secretion by inhibiting the final stages of insulin secretion with no adverse effect on more proximal events such as glucose metabolism.

To examine whether *miR-375* impairs the generation of secondary signals that are required to trigger insulin exocytosis, we measured free intracellular Ca²⁺ concentrations [Ca²⁺]_i in intact mouse pancreatic islets. Increasing the glucose concentration from 5 mM to 15 mM generated oscillations in the Ca²⁺ concentration in both the control (Fig. 3a) and the Ad-375 expressing islets (Fig. 3b). Similar oscillations were observed when the islets were stimulated with tolbutamide (Fig. 3a, b), and depolarization with high extracellular K⁺ increased [Ca²⁺]_i to the same extent in Ad-eGFP- and Ad-375-infected islets. Similar results were obtained when MIN6 cells were infected with Ad-eGFP and Ad-375 (see Supplementary Table 3). We conclude that the effects of *miR-375* on insulin secretion do not result from impaired [Ca²⁺]_i signalling.

To address whether exocytosis is impaired in cells infected with Ad-375, we applied high-resolution single-cell capacitance measurements of exocytosis to functionally identified β -cells^{14,15}.

In control β -cells (Ad-eGFP), a train of ten 500-ms depolarizations elicited an increase in membrane capacitance of 837 ± 244 fF ($1 \text{ fF} = 10^{-15} \text{ F}$; $n = 9$). In cells infected with Ad-375, the corresponding increase was limited to 94 ± 27 fF ($n = 10$; $P < 0.01$); a decrease of 85% (Fig. 3d, e). The suppression of exocytosis was not due to a decrease in Ca^{2+} entry. In both control (Ad-eGFP) and Ad-375-infected cells, the largest Ca^{2+} currents were observed during depolarizations from -70 mV to $+10 \text{ mV}$ and averaged $-50 \pm 6 \text{ pA}$ ($n = 7$) and $-48 \pm 8 \text{ pA}$ ($n = 6$), respectively (data not shown). The inhibitory action of *miR-375* remained detectable when exocytosis was induced by dialysing the cell interior with a Ca^{2+} /EGTA buffer with free Ca^{2+} concentration of $1.5 \mu\text{M}$ (Fig. 3f), thus eliciting secretion independently of Ca^{2+} influx across the plasma membrane. In these experiments, the rate of capacitance increase ($\Delta C/\Delta t$) was reduced by 63% ($P < 0.001$; $n = 15-17$) in Ad-375-infected cells compared to the control cells (Fig. 3g). In MIN6 cells, $\Delta C/\Delta t$ was reduced by $>80\%$ (Fig. 3g). The effect of *miR-375* was selective for the β -cell and no suppression of exocytosis was observed in glucagon-releasing α -cells (data not shown).

Because disruption of the actin filament network can lead to defects in exocytosis¹⁶, we also studied the effect of *miR-375* by confocal and electron microscopy. This structural analysis revealed the submembrane actin filament network to be intact (see Supplementary Fig. 3), and the granule density was unchanged in β -cells infected with either Ad-eGFP or Ad-375 (see Supplementary Figs 3 and 4). Interestingly, the reduced exocytotic capacity of the Ad-375-infected β -cells was associated with a 35% increase in the number of granules in the immediate vicinity of the plasma membrane (docked granules; see Supplementary Fig. 4). Thus, the reduced secretory capacity cannot be explained by a decreased availability of granules or by apparent defects in the submembrane actin network.

To identify genes that could mediate the observed effects on

secretion, we applied an algorithm that searches for consecutively matching base pairs between the miRNA and the target 'base-pairing nucleus' in combination with thermodynamically based evaluation of miRNA:mRNA duplex interactions¹⁷. From the compiled list of 64 putative *miR-375* target genes, we selected five genes, based on their potential role in insulin secretion and islet cell differentiation, for validation studies. These genes included: vesicle transport through interaction with t-SNAREs yeast homologue 1A (*Vti1a*)¹⁸, V-1/myotrophin (*V-1/Mtpn*)¹⁹, p38 mitogen-activated protein kinase (*Mapk14*)²⁰, monocarboxylic acid transporter member 8 (*Slc16A2*)²¹ and Max interacting protein 1 (*Mxi1*)²². The expression of these genes was studied by immunoblotting in MIN6 and N2A neuroblastoma cells (devoid of *miR-375*) that were infected with either Ad-375 or Ad-eGFP (Fig. 4a). Expression of *miR-375* in N2A cells led to reduced protein levels of Mtpn and Vti1a, whereas expression of the other genes was unaffected. Overexpression of *miR-375* in MIN6 cells using Ad-375 also decreased expression of Mtpn (Fig. 4a). Furthermore, transfection of 2'-O-me-375 increased protein levels of Mtpn but not Vti1a in MIN6 cells (Fig. 4b; data not shown). No changes were detected in mRNA levels in Ad-375-infected cells compared to controls, indicating that the regulation of target gene expression by *miR-375* is mainly post-transcriptional (Fig. 4c, d).

To test whether the predicted *miR-375* target site in the 3' untranslated region (UTR) of the Mtpn mRNA was responsible for silencing of Mtpn expression by *miR-375*, we cloned the putative 3' UTR target site downstream of a luciferase reporter gene (pRL-Mtpn-WT) and co-transfected this vector into MIN6 cells with 2'-O-me-eGFP or 2'-O-me-375. Luciferase activity of cells transfected with the 2'-O-me-375 and pRL-Mtpn was increased ~ 2 -fold relative to cells that were co-transfected with control 2'-O-me-eGFP and pRL-Mtpn (Fig. 4e). Point mutations in the nucleus of the *miR-375* target site (pRL-Mtpn-MUT), reducing the complementarity

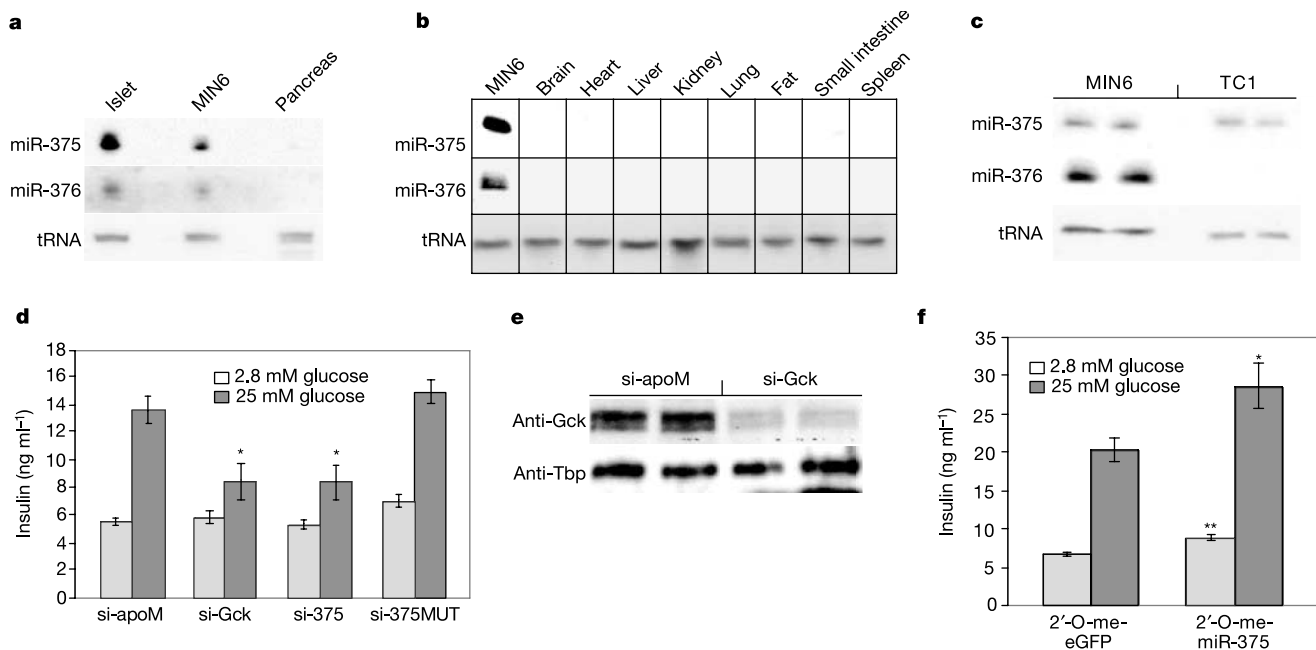


Figure 1 *miR-375* is expressed in pancreatic β -cells and regulates insulin secretion. **a**, Northern blots of total RNA ($10 \mu\text{g}$) isolated from purified pancreatic islets, MIN6 cells and total pancreas. High expression levels were detected in mouse pancreatic islets. **b**, Tissue expression of *miR-375* and *miR-376*. Total RNA ($30 \mu\text{g}$) was isolated from mouse tissues for northern blots and probed for the indicated miRNAs or transfer RNA (tRNA) as a loading control. **c**, Northern blots of total RNA ($10 \mu\text{g}$) isolated from purified MIN6 and TC1 cells. **d**, MIN6 cells were transiently transfected with synthetic siRNAs with homologous sequence to *miR-375* (si-375) or a mutated *miR-375* (si-375MUT), or siRNAs targeting *glucokinase* (si-Gck) or *apoM* (si-apoM). After 48 h, the cells were

incubated under low (2.8 mM) and stimulatory concentrations of glucose (25 mM) and insulin was measured by RIA (Linco). **e**, Immunoblot analysis of Gck in MIN6 cells that were transfected with either si-apoM (control) or si-Gck. A 70% reduction in glucokinase protein expression was observed. TATA binding protein (Tbp) was used as a loading control. **f**, MIN6 cells were transfected with 2'-O-methyl oligonucleotides complementary to *miR-375* (2'-O-me-375), or a control 2'-O-methyl oligonucleotide (2'-O-me-eGFP). Similarly, after 48 h, the cells were incubated at either 2.8 or 25 mM glucose and insulin was measured. Data represent three independent experiments $\pm$ s.e.m. with $n = 3$. * $P = 0.05$, ** $P = 0.01$.

between *miR-375* and the *Mtpn* target site, abolished the repression of endogenous *miR-375* on luciferase activity (Fig. 4e, f). These data suggest that *Mtpn* is a target of *miR-375* in pancreatic β -cells and that the repression of *Mtpn* gene expression is mediated by a single *miR-375* target site in the 3' UTR of the *Mtpn* gene.

The identification of *Vt1a* and *Mtpn* as targets for *miR-375* indicates that reduced expression of these proteins could mediate

the inhibitory action of *miR-375* on exocytosis and insulin secretion. The functions of *Vt1a* and *Mtpn* have not been studied in pancreatic β -cells, but they have been shown to be involved in vesicle transport of neurons and in neurotransmitter release^{18,23}. To test whether these proteins may contribute to the defect in secretion of Ad-375-infected cells, we silenced *Mtpn* and *Vt1a* using siRNAs in MIN6 cells and measured glucose-induced insulin secretion

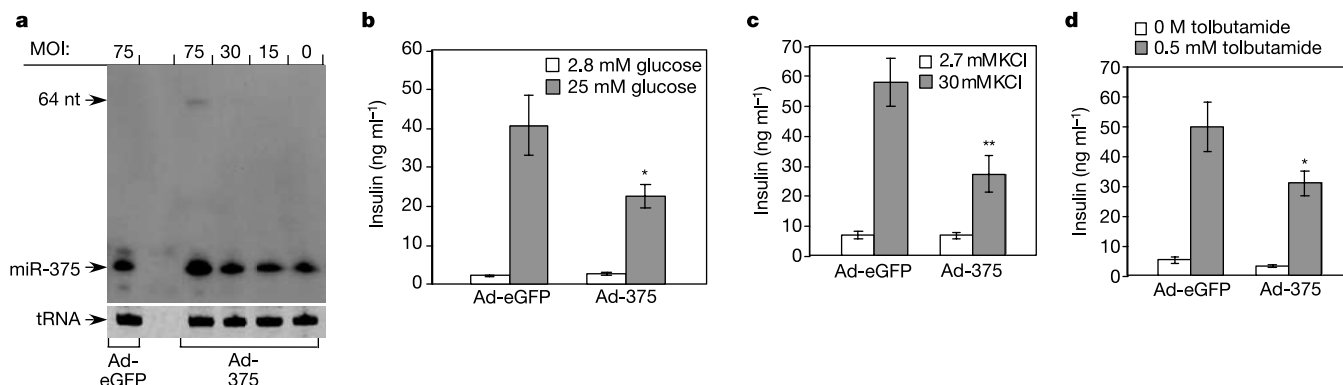


Figure 2 Expression of *miR-375* using recombinant adenovirus (Ad-375) leads to impaired glucose-, KCl- and tolbutamide-induced insulin secretion in MIN6 cells. **a**, Northern blot analysis and dose-dependent expression of *miR-375* following infection of MIN6 cells for 48 h with Ad-eGFP (control, lane 1) or Ad-375. The multiplicity of infection (MOI) is indicated. The precursor and mature *miR-375* can be visualized at ~64 and 22 nt, respectively. **b–d**, Insulin secretion of MIN6 cells following infection with

Ad-eGFP and Ad-375 in response to 25 mM glucose (**b**), 30 mM KCl (**c**) and 500 μ M tolbutamide (**d**). Insulin secretion from MIN6 cells was measured 48 h after infection with Ad-eGFP or Ad-375 and following incubation with the indicated concentrations of secretagogues. Data represent three independent experiments $\pm$ s.e.m. with $n = 3$. * $P = 0.05$, ** $P = 0.01$.

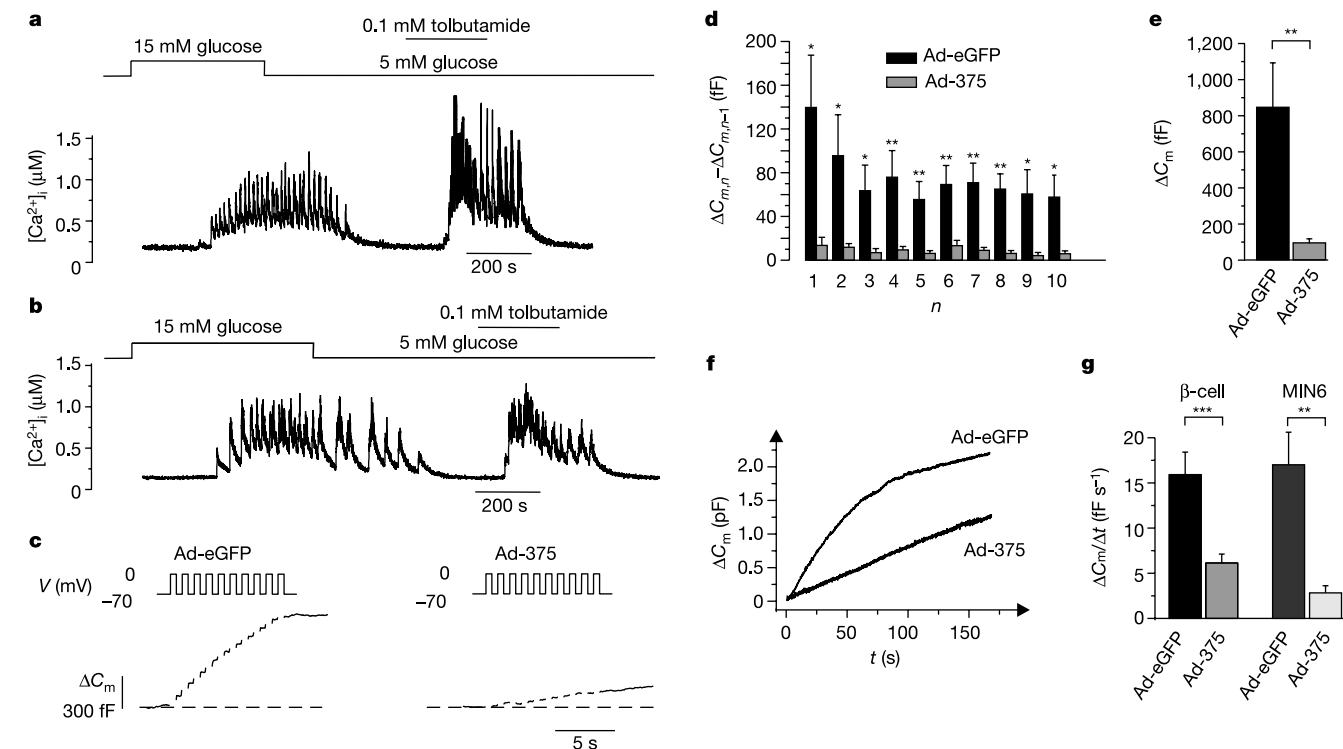


Figure 3 No effect of *miR-375* on intracellular Ca^{2+} signalling in β -cells. Intracellular $[\text{Ca}^{2+}]_i$ measurements of Ad-eGFP-infected (**a**) and Ad-375-infected pancreatic islets (**b**). The fluorescence signal has been calibrated and the approximate intracellular $[\text{Ca}^{2+}]_i$ is indicated to the left. Traces are representative of five experiments in each group. **c**, Capacitance increases (ΔC_m ; lower) elicited by a train of ten depolarizations from -70 mV to 0 mV (V; top) in β -cells infected with Ad-eGFP (left) or Ad-375 (right). **d**, Mean increase in membrane capacitance elicited by the individual depolarization of the train

($\Delta C_{m,n} - \Delta C_{m,n-1}$) displayed against pulse number (n). **e**, Total increase in membrane capacitance evoked by the train of depolarizations (ΔC_m). Data are mean values $\pm$ s.e.m. of nine or ten experiments. ** $P < 0.01$. **f**, Capacitance increase evoked by infusion of a Ca^{2+} /EGTA buffer (free $[\text{Ca}^{2+}]_i = 2 \mu\text{M}$) in one control cell and one cell over-expressing *miR-375*. **g**, Summary of the experiments in **f** and similar experiments conducted on MIN6 cells (as indicated). Data are mean values $\pm$ s.e.m. of 15–17 experiments (primary β -cells) and 7–12 measurements (MIN6 cells). ** $P < 0.01$, *** $P < 0.001$.

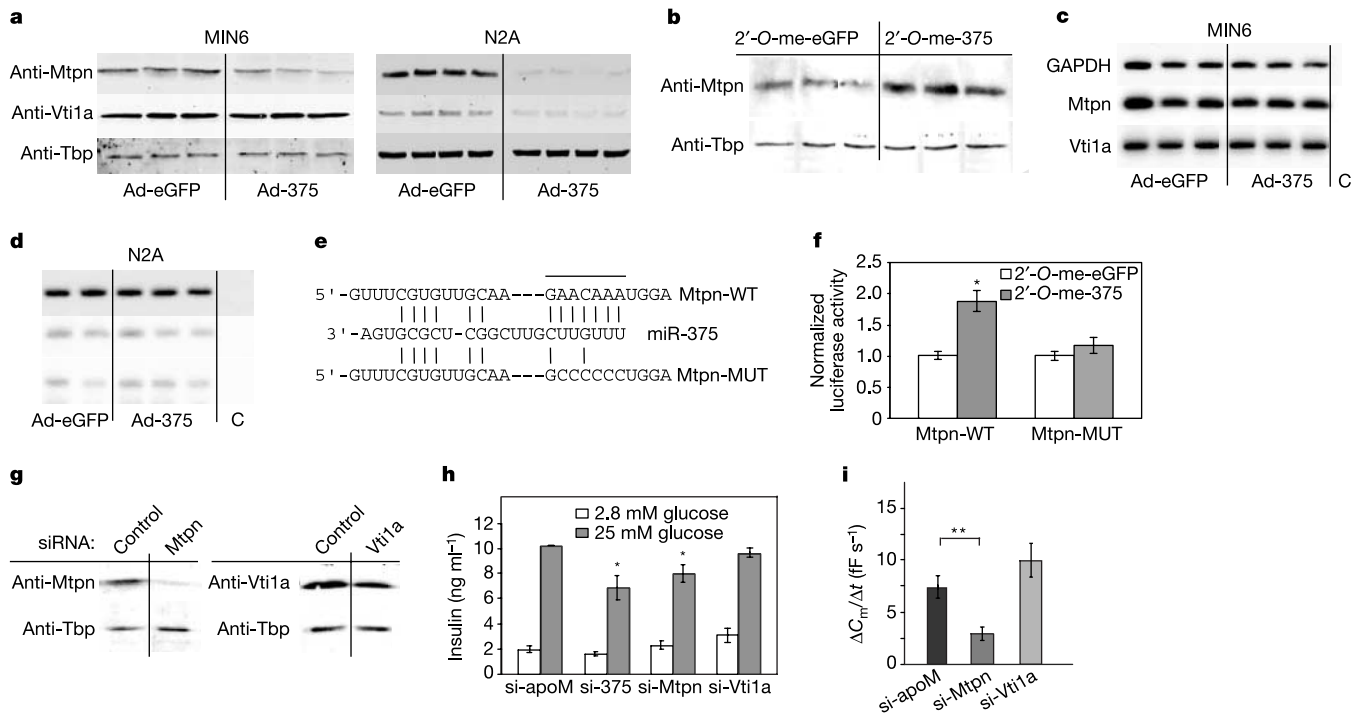


Figure 4 Identification of target genes of *miR*-375. **a**, Western blot analysis of cells infected with Ad-eGFP or Ad-375 (MIN6 cultured for 5 days post-infection; N2A, 2 days) and probed for the expression of Mtpn (anti-Mtpn), Vti1a (anti-Vti1a) or TATA binding protein (anti-Tbp) as a loading control, using specific antisera. **b**, Immunoblot analysis of Mtpn in MIN6 cells that were transfected with either 2'-O-me-eGFP (control) or 2'-O-me-375. Expression levels of TATA binding protein (Tbp) were used as a loading control; **c**, d, RT-PCR analysis of Mtpn, Vti1a and GAPDH (loading control) in MIN6 and N2A cells. **e**, Sequence of the target site in the 3' UTR of *Mtpn*. The mutant sequence (*Mtpn*-MUT) is identical to the *Mtpn*-WT construct except for five point mutations disrupting base-pairing at the 5' end of *miR*-375 (indicated with a bar). **f**, Mutating the *miR*-375 target site in the 3' UTR of *Mtpn* abolishes inhibition of luciferase activity by endogenous *miR*-375 in MIN6 cells. MIN6 cells were transiently transfected with either reporter construct in addition to 2'-O-methyl-oligonucleotides complementary to *miR*-375 (2'-O-me-375) or a control

2'-oligoribonucleotide (2'-O-me-eGFP). Data represent three independent experiments $\pm$ s.e.m. with $n = 6$. **g-i**, Silencing of *Mtpn* by siRNA impairs insulin secretion. **g**, MIN6 cells transiently transfected with siRNAs designed against *Mtpn* (si-Mtpn) or *Vti1a* (si-Vti1a) for 48 h and lysed. After separation of proteins by SDS-polyacrylamide gel electrophoresis (PAGE), samples were immunoblotted for either Mtpn or Vti1a expression. The expression of TATA binding protein (Tbp) was analysed for a loading control. **h**, MIN6 cells were transiently transfected with si-apoM (control), si-Mtpn or si-Vti1a. After 48 h, the cells were incubated under low (2.8 mM) and stimulatory concentrations of glucose (25 mM). Data represent three independent experiments $\pm$ s.e.m. with $n = 3$. $*P = 0.05$, $**P = 0.01$. **i**, Capacitance measurements in MIN6 cells transfected with si-apoM (control), si-Mtpn or si-Vti1a. Data are mean values $\pm$ s.e.m. $**P < 0.01$ versus control (si-apoM).

(Fig. 4g). Although the effect of si-Vti1a was not significant, secretion was reduced by $\sim 35\%$ in si-Mtpn-transfected cells compared to cells transfected with si-ApoM (Fig. 4h). We also verified the effects of Mtpn on insulin secretion with capacitance measurements on MIN6 cells co-transfected with si-apoM, si-Mtpn and si-Vti1a (Fig. 4i). Whereas si-Mtpn reduced exocytosis by $\sim 60\%$, decreased expression of Vti1a had no inhibitory effect. Ca^{2+} measurements on MIN6 cells co-transfected with si-apoM, si-375 and si-Mtpn showed no difference in their responses, whether they were elicited by 25 mM glucose or 30 mM K^+ (see Supplementary Table 4). Together, these results show that reduced expression of Mtpn could contribute to the defect in late stages of exocytosis induced by *miR*-375.

We have identified a functional, pancreatic islet-specific miRNA that inhibits insulin secretion at a distal stage and that occurs independently of alterations in the transmembrane Ca^{2+} fluxes and intracellular Ca^{2+} signalling. We subsequently predicted and validated *V-1/Mtpn* as a target gene of *miR*-375. The algorithm used to identify targets of *miR*-375 was based on (1) the sequence complementarity between a 3' UTR and the 5' end or 'nucleus' of *miR*-375, (2) the free energy of the miRNA:mRNA duplex, and (3) cross-species comparison of the target site. Predicted targets in mammalian cells until now have been validated using heterologous luciferase-based reporter systems²⁴. Here we have confirmed *Mtpn* as a physiological target of *miR*-375 by several independent

methods, including regulation of cellular Mtpn levels through overexpression of *miR*-375, inhibiting endogenous *miR*-375 function, and impairing exocytosis through specific silencing of *Mtpn*. However, additional targets of *miR*-375 are likely to contribute to the regulation of insulin secretion and many of the miRNAs we identified in MIN6 cells may also have roles in endocrine pancreas development. The analysis of temporal and spatial expression and the generation of loss-of-function mutations of islet miRNAs will shed light on this new class of genes in these processes. In conclusion, tissue-specific miRNAs, such as *miR*-375, have the potential to become novel targets for therapeutic intervention in diabetes mellitus. □

Methods

miRNA cloning and northern blotting analysis

Total RNA isolated from MIN6 (600 μg) or TC1 (300 μg) cells was separated on a 15% denaturing polyacrylamide gel and 19–24-nt small RNAs were recovered from the gel and used as input for adaptor ligation. Adaptor ligation and reverse transcription polymerase chain reaction (RT-PCR) of the ligation product was performed as described¹⁰. Modifications for the TC1 library can be found in the Supplementary Methods. Antisense probes were designed to complement cloned miRNA sequences²⁵.

Cell culture

MIN6 cells were cultured with DMEM medium containing 25 mM glucose, 15% fetal bovine serum (FBS) and 5.5 μM 2-mercaptoethanol. N2A cells were cultured with DMEM medium containing 25 mM glucose and 10% FBS.

Insulin secretion studies

Insulin secretion in MIN6 cells was performed as previously described²⁶ by RIA (Linco Research). Insulin content of MIN6 cells and pancreatic islets was measured as previously described²⁷.

Generation of recombinant adenovirus

The recombinant adenovirus used to express *miR-375* (Ad-375) was generated by PCR, amplifying the miRNA precursor sequence with primers: 5'-CCCCAAGGCTGATGCTGAGAAGCCGCCCC-3' and 5'-GCCGCCCGGCCCGGGTCTTC-3'. The fragment was inserted into shuttle vector Ad5CMV-K NpA. Ad-eGFP (ViraQuest) does not contain a transgene, and was used for control. Cells or islets were infected at an MOI of 25–50 viral particles per cell in DMEM with 2% FBS and cultured for 48 h prior to experimentation.

Electrophysiology and Ca²⁺ measurements

Measurements of exocytosis and inward Ca²⁺ currents were conducted on single mouse β -cells or MIN6-cells ~24 h after infection with Ad-eGFP or Ad-375, or after transfection with siRNAs directed against *ApoM*, *Mtpn* and *Vt1a*. The electrophysiological recordings were analysed by standard whole-cell configuration of the patch-clamp technique as described previously²⁸. The identity of the α - and β -cells was established as described previously¹⁴. [Ca²⁺]_i was measured by dual-excitation wavelength spectrofluorimetry²⁹. All electrophysiological experiments and Ca²⁺ measurements were carried out at 32–34 °C. The infection of islets and loading with the Ca²⁺ indicator were evaluated using a Zeiss LSM510 microscope (Carl Zeiss). eGFP was excited at 488 nm, whereas Fura-2 was excited at 820 nm (using two-photon excitation) line. Emitted light was visualized using a $\times 40$ water objective and separated using the hardware and software of the META package (Carl Zeiss).

Assay of luciferase activity

The mouse myotrophin 3' UTR target site was PCR-amplified using the following primers: 5'-TCCATCATTTTCATATGCACTGTATC-3' and 5'-TCATATCGTTAAGGACGCTCGGAAA-3' and cloned downstream of the stop codon in pRL-TK (Promega). This construct was used to generate the mutant myotrophin plasmid (Fig. 4d). MIN6 cells were cultured in 24-well plates and each transfected with 0.4 μ g of pRL-TK (*Rr-luc*) and 0.1 μ g of pGL3 control vector (*Pp-luc*) (Promega). Cells were harvested and assayed 30–36 h post-transfection.

Electron microscopy, immunocytochemistry and live cell imaging

Electron microscopy was performed as described²⁹, except islets were embedded in Durcupan (Sigma). Prior to fixation, infection of the peripheral cells was ascertained by confocal microscopy, and subsequent electron microscopic analyses were restricted to cells in the islet periphery (see Supplementary Fig. 1). The distribution of actin in mouse β -cells and MIN6 cells infected with Ad-eGFP or Ad-375 was analysed using AlexaFluor 532-phalloidin (Molecular Probes). The fluorescence was detected using a Zeiss Pascal microscope at $\times 100$ objective and excitation lines 488 and 543 nm to detect eGFP and Alexa532, respectively. Emitted light was collected at >560 nm (Alexa532) and 505–530 nm (eGFP).

Identification of miR-375 targets

To identify targets of *miR-375*, we used our recently developed algorithm as described¹⁷. The core algorithm consists of two steps: (1) the search for a GC-rich string of consecutive complementary bases ('nucleus') between the miRNA and the putative target sequence in the 3' UTRs of mRNAs, and (2) *in silico* evaluation of the free energy of the predicted miRNA:mRNA duplexes³⁰. The algorithm was applied to the Refseq data set (Release 1, 14 April 2003, ftp://ftp.ncbi.nih.gov/refseq/). A more detailed description of the analysis can be found in the Supplementary Methods.

siRNA and 2'-O-methyl oligoribonucleotides

Synthetic miRNAs and siRNAs were synthesized by Dharmacon. siRNA SMARTPOOLS (mixtures of four unique siRNA duplexes) were designed from the mouse myotrophin (GenBank accession number NM_008098) and mouse *Vt1a* (NM_016862) sequences. The sequence of si-375MUT is TTTGAAGGTTTCGGCTCGCGTT, 2'-O-me-eGFP is AAGGCAAGCUGACCCUGAAGUL, and 2'-O-me-375 is UGCAUCACGCGAGCCGAA CGAACAAUAAGL. All 2'-O-methyl oligonucleotides were synthesized as previously described¹². Reagents were either transfected into MIN6 cells using Lipofectamine 2000 (Invitrogen) at 200 nM, or 5 μ g were electroporated using the Amaxa Nucleofector system.

Antibodies

Antibodies for immunoblotting were obtained from different sources: anti-myotrophin (gift of M. Taoka), anti-Vt1a (BD Transduction Laboratories), anti-p38 MAPK (Cell Signaling), anti-MCT8 (gift of A. Halestrap), anti-Mxl1 (Santa Cruz) and anti-TATA box binding protein (gift of R. Roeder).

RT-PCR

Extraction of total RNA, synthesis of cDNA, and PCR were carried out as previously described²⁷. Primer sequences used for PCR are available upon request.

Cellular ATP measurements

Intracellular ATP was measured using the Bioluminescent Somatic Cell Assay Kit (Sigma) according to the manufacturer's instructions.

Statistical analysis

Results are given as mean $\pm$ s.d. Statistical analyses were performed by using Student's *t*-test, and the null hypothesis was rejected at the 0.05 level.

Received 5 August; accepted 30 September 2004; doi:10.1038/nature03076.

- Bartel, D. P. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* **116**, 281–297 (2004).
- Lagos-Quintana, M., Rauhut, R., Lendeckel, W. & Tuschl, T. Identification of novel genes coding for small expressed RNAs. *Science* **294**, 853–858 (2001).
- Lau, N. C., Lim, L. P., Weinstein, E. G. & Bartel, D. P. An abundant class of tiny RNAs with probable regulatory roles in *Caenorhabditis elegans*. *Science* **294**, 858–862 (2001).
- Lee, R. C. & Ambros, V. An extensive class of small RNAs in *Caenorhabditis elegans*. *Science* **294**, 862–864 (2001).
- Carrington, J. C. & Ambros, V. Role of microRNAs in plant and animal development. *Science* **301**, 336–338 (2003).
- Wightman, B., Ha, I. & Ruvkun, G. Posttranscriptional regulation of the heterochronic gene *lin-14* by *lin-4* mediates temporal pattern formation in *C. elegans*. *Cell* **75**, 855–862 (1993).
- Olsen, P. H. & Ambros, V. The *lin-4* regulatory RNA controls developmental timing in *Caenorhabditis elegans* by blocking LIN-14 protein synthesis after the initiation of translation. *Dev. Biol.* **216**, 671–680 (1999).
- Brennecke, J., Hipfner, D. R., Stark, A., Russell, R. B. & Cohen, S. M. bantam encodes a developmentally regulated microRNA that controls cell proliferation and regulates the proapoptotic gene *hid* in *Drosophila*. *Cell* **113**, 25–36 (2003).
- Xu, P., Vernooy, S. Y., Guo, M. & Hay, B. A. The *Drosophila* microRNA Mir-14 suppresses cell death and is required for normal fat metabolism. *Curr. Biol.* **13**, 790–795 (2003).
- Pfeffer, S., Lagos-Quintana, M. & Tuschl, T. In *Current Protocols in Molecular Biology* (eds Ausubel, F. M. B. R. *et al.*) Ch. 26.4.1–26.4.18 (Wiley Interscience, New York, 2003).
- Ambros, V. *et al.* A uniform system for microRNA annotation. *RNA* **9**, 277–279 (2003).
- Meister, G., Landthaler, M., Dorsett, Y. & Tuschl, T. Sequence-specific inhibition of microRNA- and siRNA-induced RNA silencing. *RNA* **10**, 544–550 (2004).
- Hutvagner, G., Simard, M. J., Mello, C. C. & Zamore, P. D. Sequence-specific inhibition of small RNA function. *PLoS Biol* **2**, E98 (2004).
- Barg, S., Galvanovskis, J., Gopel, S. O., Rorsman, P. & Eliasson, L. Tight coupling between electrical activity and exocytosis in mouse glucagon-secreting alpha-cells. *Diabetes* **49**, 1500–1510 (2000).
- Gopel, S. O., Kanno, T., Barg, S. & Rorsman, P. Patch-clamp characterisation of omatostatin-secreting cells in intact mouse pancreatic islets. *J. Physiol.* **528**, 497–507 (2000).
- Tsuboi, T., da Silva Xavier, G., Leclerc, I. & Rutter, G. A. 5'-AMP-activated protein kinase controls insulin-containing secretory vesicle dynamics. *J. Biol. Chem.* **278**, 52042–52051 (2003).
- Rajewsky, N. & Sock, N. D. Computational identification of microRNA targets. *Dev. Biol.* **267**, 529–535 (2004).
- Antonin, W., Riedel, D. & von Mollard, G. F. The SNARE *Vt1a*- β is localized to small synaptic vesicles and participates in a novel SNARE complex. *J. Neurosci.* **20**, 5724–5732 (2000).
- Taoka, M. *et al.* V-1, a protein expressed transiently during murine cerebellar development, regulates actin polymerization via interaction with capping protein. *J. Biol. Chem.* **278**, 5864–5870 (2003).
- Larsen, C. M. *et al.* Interleukin-1 β -induced rat pancreatic islet nitric oxide synthesis requires both the p38 and extracellular signal-regulated kinase 1/2 mitogen-activated protein kinases. *J. Biol. Chem.* **273**, 15294–15300 (1998).
- Zhao, C., Wilson, M. C., Schuit, F., Halestrap, A. P. & Rutter, G. A. Expression and distribution of lactate/monocarboxylate transporter isoforms in pancreatic islets and the exocrine pancreas. *Diabetes* **50**, 361–366 (2001).
- Schreiber-Agus, N. *et al.* An amino-terminal domain of Mx1 mediates anti-Myc oncogenic activity and interacts with a homolog of the yeast transcriptional repressor SIN3. *Cell* **80**, 777–786 (1995).
- Yamakuni, T. *et al.* V-1, a catecholamine biosynthesis regulatory protein, positively controls catecholamine secretion in PC12D cells. *FEBS Lett.* **530**, 94–98 (2002).
- Lewis, B. P., Shih, I. H., Jones-Rhoades, M. W., Bartel, D. P. & Burge, C. B. Prediction of mammalian microRNA targets. *Cell* **115**, 787–798 (2003).
- Lagos-Quintana, M. *et al.* Identification of tissue-specific microRNAs from mouse. *Curr. Biol.* **12**, 735–739 (2002).
- Minami, K. *et al.* Insulin secretion and differential gene expression in glucose-responsive and -unresponsive MIN6 sublines. *Am. J. Physiol. Endocrinol. Metab.* **279**, E773–E781 (2000).
- Shih, D. Q. *et al.* Profound defects in pancreatic beta-cell function in mice with combined heterozygous mutations in Pdx-1, Hnf-1 α , and Hnf-3 β . *Proc. Natl Acad. Sci. USA* **99**, 3818–3823 (2002).
- Eliasson, L. *et al.* SUR1 regulates PKA-independent cAMP-induced granule priming in mouse pancreatic B-cells. *J. Gen. Physiol.* **121**, 181–197 (2003).
- Olofsson, C. S. *et al.* Fast insulin secretion reflects exocytosis of docked granules in mouse pancreatic B-cells. *Pflügers Arch.* **444**, 43–51 (2002).
- Zuker, M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* **31**, 3406–3415 (2003).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank K. Borglid, J. Chen, J. Galvanovskis, M. Lagos-Quintana, M. Landthaler, A. Lingqvist, G. Meister, B. M. Nilsson, A. Wendt and C. Wolfrum for advice and technical assistance. This work was supported by an unrestricted grant from Bristol Myers Squibb, the Juvenile Diabetes Research Foundation, the Deutsche Forschungsgemeinschaft and grants from the Swedish Research Council, the Swedish Diabetes Association, the Göran Gustafsson Stiftelse for Natural Sciences and Medicine and the Swedish Strategic Research Foundation (SSF).

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to M.S. (stoffel@rockefeller.edu).

Processing of primary microRNAs by the Microprocessor complex

Ahmet M. Denli¹, Bastiaan B. J. Tops², Ronald H. A. Plasterk², René F. Ketting² & Gregory J. Hannon¹

¹Cold Spring Harbor Laboratory, Watson School of Biological Sciences, 1 Bungtown Road, Cold Spring Harbor, New York 11724, USA

²The Hubrecht Laboratory Centre for Biomedical Genetics, Uppsalalaan 8, 3584 CT Utrecht, The Netherlands

Mature microRNAs (miRNAs) are generated via a two-step processing pathway to yield ~22-nucleotide small RNAs that regulate gene expression at the post-transcriptional level¹. Initial cleavage is catalysed by Drosha, a nuclease of the RNase III family, which acts on primary miRNA transcripts (pri-miRNAs) in the nucleus². Here we show that Drosha exists in a multi-protein complex, the Microprocessor, and begin the process of deconstructing that complex into its constituent components. Along with Drosha, the Microprocessor also contains Pasha (partner of Drosha), a double-stranded RNA binding protein. Suppression of Pasha expression in *Drosophila* cells or *Caenorhabditis elegans* interferes with pri-miRNA processing, leading to an accumulation of pri-miRNAs and a reduction in mature miRNAs. Finally, depletion or mutation of *pash-1* in *C. elegans* causes de-repression of a *let-7* reporter and the appearance of phenotypic defects overlapping those observed upon examination of worms with lesions in Dicer (*dcr-1*) or Drosha (*drsh-1*). Considered together, these results indicate a role for Pasha in miRNA maturation and miRNA-mediated gene regulation.

miRNAs are a class of small, non-coding RNAs that enter the RNA interference (RNAi) pathway to regulate the expression of protein-encoding genes at the post-transcriptional level³. miRNA production begins with the synthesis of pri-miRNAs, ranging in size from several hundred nucleotides (nt) to several kilobases. These are recognized and cleaved into precursor miRNAs (pre-miRNAs) in the nucleus by an RNase III family nuclease, Drosha²; the pre-

miRNAs are short, hairpin RNAs of approximately 70 nt, bearing the 2-nucleotide 3' overhang that is a signature of RNase III-mediated cleavage. Pre-miRNAs are exported to the cytoplasm by a RanGTP/exportin 5-dependent mechanism^{4–6}, with their characteristic overhang contributing to their entry into this export pathway⁵. Once in the cytoplasm, pre-miRNAs are recognized and processed into their mature, ~22-nt form by Dicer^{7–9}, with the 3' overhang again playing a role in specifying cleavage^{10–12}. Mature miRNAs enter RISC (RNA-induced silencing complex) in an asymmetric fashion such that for most miRNAs, only one strand is enabled to recognize and repress the expression of target genes¹³. The outcome of this recognition is either endonucleolytic cleavage of the targeted messenger RNA^{14,15} or interference with protein synthesis by a mechanism that remains unclear^{16–18}. With the ultimate goal of addressing the mechanisms by which pri-miRNAs are tagged for entry into the RNAi pathway and processed at specific sites, we have undertaken a biochemical characterization of Drosha and its associated factors.

Previously, Kim and colleagues showed² that epitope-tagged, human Drosha protein was capable of releasing pre-miRNAs from pri-miRNAs *in vitro* and contributed to miRNA maturation *in vivo*. Processing was dependent upon the presence of a double-stranded region around the cleavage position. We asked whether a pri-miRNA processing activity also existed in *Drosophila* cells. Although they are absent from *Schizosaccharomyces pombe* and *Arabidopsis*, homologues of mammalian Drosha are present in *Drosophila* and *C. elegans*¹⁹. Indeed, S2 cell extracts contain an activity that can recognize a pri-miRNA, in this case pri-bantam, and cleave this into discrete products that can be identified as pre-bantam and the regions of the pri-miRNA that flank that mature sequence; pre-bantam and the 5' and 3' flanks are 60, 104 and 112 nt, respectively (Fig. 1a; see Supplementary Fig. S1 for sequences). Immunoprecipitates obtained using an affinity-purified Drosha antibody were capable of generating pre-bantam from pri-bantam (Fig. 1b). Examination of the supernatants showed that a substantial fraction of pri-miRNA processing activity was depleted from the extract by immunoprecipitation (data not shown). Additional pri-miRNAs from both human and *Drosophila* were similarly processed by extracts and immunoprecipitates (data not shown). Considered together, these results implicate Drosha and/or

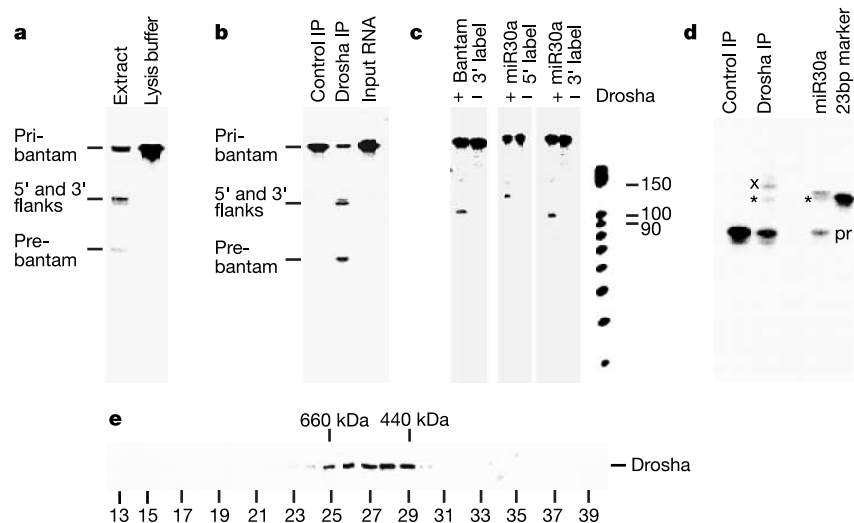


Figure 1 Pri-miRNA processing is conserved in *Drosophila*. **a**, Pri-miRNA processing activity was assessed in *Drosophila* S2 cell extracts using uniformly labelled pri-bantam miRNA. Positions of the processed pre-bantam and the two flanking RNAs derived from excision of pre-bantam are indicated. **b**, Immunoprecipitation experiments with S2 extracts test whether processing activity associates with Drosha. **c**, To assess the accuracy of pri-miRNA processing, cleavage sites were mapped using end-labelled substrates. Pre-bantam or pre-miR30a was 3' labelled by ligation to ³²P-pCp, and

pri-miR30a was 5' labelled by capping with ³²P-GTP. Sizes of cleavage products were as predicted (see Supplementary Fig. S1). RNA size markers, indicated in bases, are shown at the right. **d**, The 5' end of processed pre-miR30a was mapped by primer extension. Control extension of a synthetic pre-miR30a RNA is shown for reference along with a 23 bp 5'-end-labelled DNA marker. **e**, Western blots with anti-Drosha antibody across fractions of a Superose 6 column show that Drosha from S2 cells exists in a ~500 kDa complex. Fraction numbers are given below the panel.

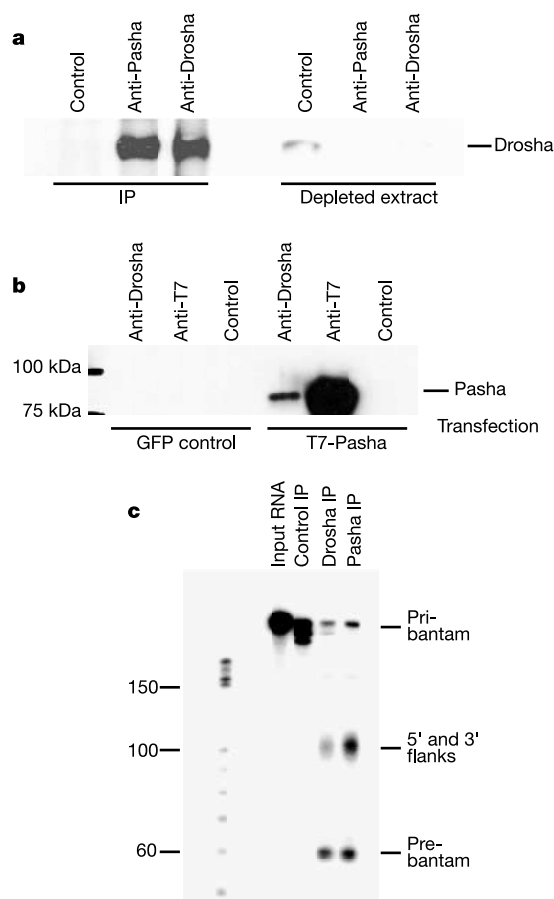


Figure 2 Pasha, a nuclear dsRNA binding domain (dsRBD) protein, associates with Drosha. **a**, Immunoprecipitates from S2 cell extracts with antibodies directed against endogenous Drosha or Pasha proteins were examined by western blotting with anti-Drosha antibody. To assess the degree of co-association, supernatants were also probed. The immunoprecipitates were derived from extract equivalent to $\sim 40 \times$ of what was loaded in the supernatant lanes. **b**, Immunoprecipitations were prepared from S2 cells that were transfected with either a GFP expression construct or a construct expressing T7 epitope-tagged Pasha. **c**, Immunoprecipitates were incubated with uniformly labelled pri-bantam and tested for processing activity.

its associated factors as the major source of pri-miRNA processing activity in *Drosophila* extracts.

The accuracy of pri-miRNA processing by our immunoaffinity-purified Drosha was assessed in two ways. First, we examined the processing of two pri-miRNA substrates that were labelled at their 5' or 3' termini (Fig. 1c). Cleavage of pri-bantam or pri-miR30a yielded the expected end fragments. The end of pre-miR30a has previously been mapped by primer extension of *in vivo* and *in vitro* processed transcripts¹. Upon analysis of pri-miR30a, processed using immunoaffinity-purified *Drosophila* Drosha, we find two prominent extension products. The smaller (denoted with an asterisk in Fig. 1d) is consistent with the expected pre-miR30a product, based both on a labelled DNA marker and on primer extension of a synthetic version of the expected pre-miR30a product. A larger product (denoted with 'X' in Fig. 1d), differing from the expected product by 2–3 nt, is of unknown origin. However, it should be noted that many nucleases, including Dicer, are known to make multiple cleavages within their substrates. In contrast, Drosha does not operate on fully duplexed substrates²⁰ (Supplementary Fig. S2).

To investigate how Drosha selects its substrates and how the enzyme determines its specific cleavage sites, we sought to characterize native complexes containing the enzyme. Biochemical fractionation indicates that Drosha is present within an ~ 500 kDa complex, which presumably contains additional protein components (Fig. 1e). Given its role in microRNA metabolism, we have dubbed this complex the Microprocessor.

Recent, genome-wide two-hybrid analysis of *Drosophila* has generated a substantial list of candidate protein–protein interactions²¹. In examining the list, we noted a potential interaction between Drosha and a double-stranded RNA (dsRNA) binding protein, originally dubbed CG1800 (candidate gene 1800). This protein consists of two domains, a WW domain near the amino terminus and a dsRNA binding domain near the carboxyl terminus. Notably, WW domains interact with proline-rich domains, such as the one present near the N terminus of Drosha. Along with Drosha, CG1800 is conserved in *C. elegans* and mammals, where it was called DGCR8²². Like Drosha, CG1800 homologues are absent from the *Arabidopsis* and *S. pombe* genomes. Results from the published two-hybrid studies, as well as evidence of physical and genetic interaction presented below, caused us to designate CG1800 as Pasha (partner of Drosha).

Consistent with the possibility that they might both be present in a single complex, Drosha and Pasha are predominantly nuclear

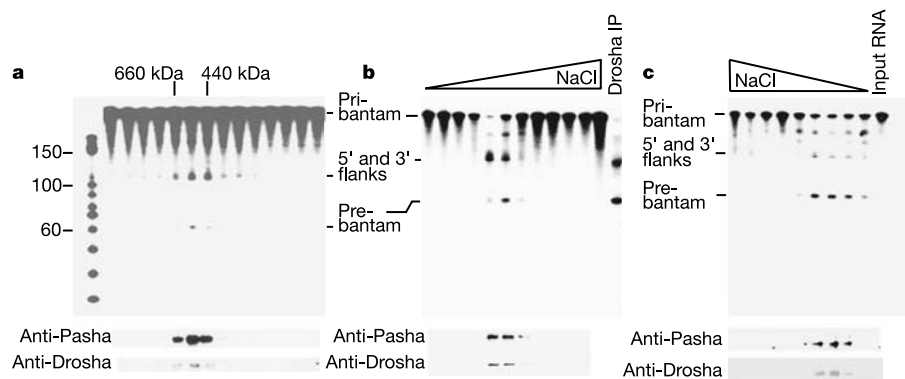


Figure 3 Drosha and Pasha co-exist in a Microprocessor complex. **a**, Combined nuclear and cytoplasmic extracts from S2 cells were loaded onto a Superose 6 column, and fractions were subjected to western blotting with Drosha or Pasha antibodies (lower panel). The Drosha antibody was used to recover Drosha complexes from each fraction and these were tested for their ability to process pri-bantam (upper panel). Positions of pre-bantam and the 5' and 3' flanks are indicated. **b**, Extracts prepared as in **a** were run on a Q sepharose FF column. Drosha and Pasha fractionation profiles were determined by

western blotting (lower panels). Pri-miRNA processing activity on pri-bantam was examined using Drosha immunoprecipitates from each fraction. Direct analysis of processing activity in each fraction showed a similar processing activity profile, albeit with weaker activity than exhibited by the immunoprecipitates. **c**, Drosha peak fractions from a Q-sepharose column were pooled and loaded onto a phenyl HP column. Fractionation of Drosha and Pasha and processing activity were followed as in **b**.

proteins (Supplementary Fig. S3; data not shown). Furthermore, Pasha antiserum co-immunoprecipitates Drosha, essentially depleting it from whole cell extracts (Fig. 2a). This interaction was not disrupted by treatment of extracts or immunoprecipitates with RNase (data not shown). Analysis of the reciprocal immunoprecipitate required the use of a T7-tagged Pasha, since the antibody heavy chain interferes with detection of this protein in western blots

with the rabbit anti-Pasha antibody. In T7-Pasha-expressing cells, a Drosha antibody co-immunoprecipitates tagged Pasha (Fig. 2b). Drosha-Pasha complexes are functional since both Drosha and Pasha immunoprecipitates are able to process pri-miRNA into pre-miRNA *in vitro* with similar efficiencies (Fig. 2c). To test whether Drosha and Pasha coexist in the 500 kDa Microprocessor, we examined chromatographic profiles from S2 cell extracts. Both

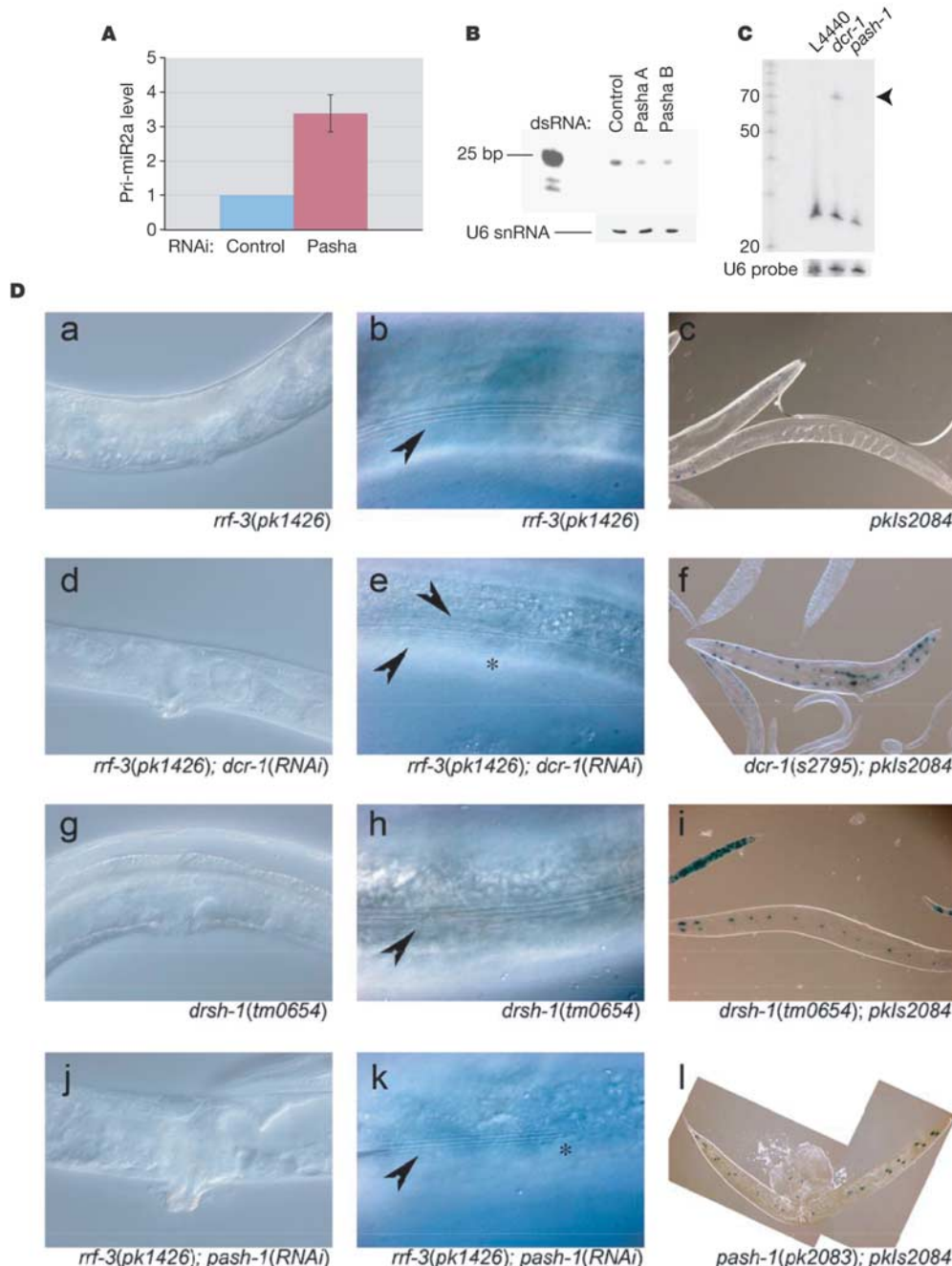


Figure 4 Depletion of Pasha reduces pri-miRNA processing. **A**, Pri-miR2a levels were analysed following transfection of *Drosophila* S2 cells with two different dsRNAs homologous to *Pasha* by semiquantitative RT-PCR. Error bar indicates standard deviation from the mean. **B**, Total RNA extracted from S2 cells transfected with either a control dsRNA (luc) or either of two Pasha dsRNAs was examined for the level of mature miR2a by northern blotting. U6 snRNA was used as a control to ensure equal loading. **C**, *C. elegans* (*rrf-3* mutant) were treated with Pasha (*pash-1*), Dicer (*dcr-1*) or control dsRNAs (L4440), and total RNA was prepared. Levels of pre- and mature *let-7* were assessed by northern blotting. Sizes of selected RNA markers are indicated in bases at the left. **D**, Phenotypic analysis of Pasha (*pash-1*) and Drosha (*drsh-1*) in *C. elegans* was performed using RNAi and genetic

mutants. RNAi against *pash-1* leads to vulva (100%) and alae (10/17) defects, similar to those published for *dcr-1* mutants^{8,9} (**a** and **b**, **d** and **e**, **g** and **h**, and **j** and **k**). The observed defects include protrusion and bursting of the vulva, and gaps in, or absence of the alae (asterisks). Occasionally duplicated alae are detected, as shown in **e**. The *drsh-1(tm0654)* allele does not result in obvious defects in these structures (alae indicated by arrow heads); however this mutant may be partially rescued by maternal contribution of Drosha. **i** and **l**, both *drsh-1(tm0654)* (10 out of 10) and *pash-1(pk2083)* (14 out of 14) result in defects in the *let-7*-mediated silencing of a previously described transgene (*pkl-2084*)²⁴. This defect is similar to that observed in a *dcr-1(s2795)* mutant (**f**: 15 out of 15). In wild-type adults, the *lacZ* expression is silenced (**c**: 0 out of 57 showing *lacZ* expression).

proteins along with pri-miRNA processing activity peaked in an ~500 kDa fraction (Fig. 3a). Similar co-fractionation was also seen if complexes were serially chromatographed on Q sepharose FF and phenyl HP columns (Fig. 3b, c) or on Q sepharose FF and mono S (data not shown).

To examine the functional connection between Pasha and miRNA metabolism, we asked whether suppression or mutation of Pasha had any effect on pri-miRNA processing or miRNA function. In *Drosophila* S2 cells, Pasha dsRNAs caused a modest suppression of Pasha mRNA (Supplemental Fig. S4). This resulted in both an accumulation of pri-miR2a and a reduction in levels of mature miR2a (Fig. 4a, b). Similar effects on pri-miR30a were seen upon transfection of mammalian cells with two different human Pasha siRNAs (Supplementary Fig. S5). Similarly, treatment of *C. elegans* with *pash-1* dsRNA, by feeding of RNAi-hypersensitive *rrf-3* mutant worms, resulted in an accumulation of pri-*let-7* with a decrease in the mature species (Fig. 4c; Supplementary Fig. S6). Notably, similar effects were seen upon suppression of Dicer (*dcr-1*), although in this case the pre-miRNA also accumulated. A comparison of these two results suggests that Pasha acts in pri-miRNA metabolism upstream of pre-miRNA production, consistent with its placement by biochemistry as a component of the Microprocessor. Precisely why pri-miRNAs also accumulate in worms treated with Dicer dsRNA but not in worms treated with control dsRNAs is at present unclear.

For studies of the biological function of Drosha and Pasha *in vivo*, we again turned to *C. elegans*. A deletion allele of Drosha, *drsh-1(tm0654)* causes a sterile phenotype, with no other visible defects. Specifically, we looked at the presence of alae and the structure of the vulva (Fig. 4D, g and h), as these structures are often affected by lesions in miRNA pathway genes like *dcr-1* (Fig. 4D, d and e)^{8,9}. Most probably, such defects are not observed in *drsh-1(tm0654)* animals because of a strong maternal rescue. Defects in the alae and vulva structures are easily detected in *pash-1* RNAi knock-down animals and to a lesser degree in a *pash-1(pk2083)* nonsense mutant (Fig. 4D, j and k; data not shown). Typical defects include protrusion or bursting of the vulva, and gaps or absence of the alae. In addition, we used a more sensitive and specific assay to directly monitor the activity of the *let-7* miRNA. The assay is based on a *lacZ* reporter that is silenced by *let-7* through sequences in the 3' UTR²³. This results in detectable *lacZ* staining in all larval stages because *let-7* is not yet expressed in these stages, but an absence of *lacZ* in the *let-7* expressing adult (Fig. 4D, c). As a control, we used *dcr-1* mutant animals. Indeed, in this mutant background *lacZ* is reactivated in the adult (Fig. 4D, f)²⁴. We then asked whether the adult-specific silencing of this reporter requires *drsh-1* and/or *pash-1*. Both *pash-1(pk2083)* and *drsh-1(tm0654)* lead to re-expression of the reporter in the adult stage (Fig. 4D, l and i), indicating that *in vivo*, both Drosha and Pasha proteins are required for the function or synthesis of the *let-7* miRNA.

Considered together, our results indicate that Pasha and Drosha are components of a multiprotein machine, the Microprocessor, which converts pri-miRNAs into pre-miRNAs. Although the experiments presented here do not permit us to assign a definitive function to Pasha within the Microprocessor, it is reasonable to speculate that Pasha might have one of several roles. For example, it could help in identifying primary miRNA transcripts, facilitating delivery of these to Drosha for cleavage. Indeed, Pasha immunoprecipitates contain pri-miRNAs (Supplementary Fig. S7). Alternatively, Pasha could help to orient the pri-miRNA in the Microprocessor, contributing to the specific positioning of the Drosha cleavage site. A parallel study shows an association between DGCR8 and human Drosha, and presents biochemical and genetic evidence that these proteins cooperate to determine the specificity of Drosha cleavage²⁵. On the basis of the results presented here, Pasha joins RDE-4, Hyl-1 and R2D2^{26–28} to form a growing list of dsRNA-binding proteins that play important yet distinct roles in the RNAi pathway. □

Methods

Drosophila cell culture and extract preparation

Drosophila S2 cells were cultured as described previously²⁹. Extracts were prepared in two different ways. The first method was mainly used for chromatographic studies. Cells were washed with cold PBS and then lysed in buffer 5A (10 mM HEPES pH 7.0, 10 mM NaCl, 1.5 mM MgCl₂, 0.75 mM DTT, 0.2 mM PMSF, protease inhibitors (Roche)) for 30 min and dounce homogenized. Nuclei were spun down, and the supernatant was saved as cytoplasmic extract. The nuclei were resuspended in buffer 5B (10 mM Tris pH 8.0, 500 mM NaCl, 1.5 mM MgCl₂, 20% glycerol, 0.75 mM DTT, 0.2 mM PMSF, protease inhibitors (Roche)) and extracted for 40 min on ice. The extracted nuclei were spun down and the supernatant saved as the nuclear fraction. Nuclear and cytoplasmic fractions were combined and spun at 20,000 g for 30 min (S10). The combined extract was used in most experiments. The second method was principally for immunoprecipitations and involved lysis of cells in buffer DmLB10 (25 mM Tris pH 7.5, 150 mM NaCl, 1.5 mM MgCl₂, 1% Triton X-100, 1 mM DTT, 0.2 mM PMSF, protease inhibitors (Roche)) and a 20,000 g 30 min spin.

Immunofluorescence

A modification of the protocol in ref. 30 was used. S2 cells were grown on slides coated with Concanavalin A (Sigma).

Pri-miRNA processing assay

DNA templates for transcription of Drosha substrates were reverse transcribed with the Superscript II Kit (Invitrogen) from total RNA prepared with Trizol (Invitrogen) using primers that flank the pre-miRNAs at least 50 bases. Primer sequences as follows: mir30a fwd and rev (see ref. 1); mir2a-1 T7 fwd: 5'-TAATACGACTCACTAGGGATTCTGAACTCCCCGAGTCAAGTATCC-3', mir2a-1 rev: 5'-GTGTGAATTATGTGGCGGGAGGTTATT-3'; mir-bantam T7-fwd: 5'-TAATACGACTCACTAGGGCGCTCAGATGCGAGATGTTGTGTGAT-3', mir-bantam rev: 5'-GATCGGTGGCATAAGTTCAAAGC-3'. Pri-miRNAs were transcribed and uniformly labelled with ³²P (Perkin Elmer) using the Riboprobe Kit (Promega). 5' and 3' labelling of cold-transcribed (Megascript kit, Ambion) RNA was done using guanylyltransferase (Ambion) and T4 DNA ligase (Amersham), respectively. A typical pri-miRNA processing reaction contained 1 µl 10xDT (100 mM Tris pH 8.0, 60 mM MgCl₂, 5 mM DTT), 1 µl Rnasin (Promega), 0.2 µl ³²P-labelled pri-miRNA, 5 µl extract, 2.8 µl H₂O. Reactions were allowed to proceed for 2 h at 30 °C.

Chromatography

AKTA FPLC was used for column chromatography. All the columns and media were purchased from Amersham Biosciences. Buffers for ion exchange chromatography: IEX'A': 10 mM Tris pH 7.5, 100 mM NaCl, 1.5 mM MgCl₂, 4% glycerol, 0.75 mM DTT; IEX'B': 10 mM Tris pH 7.5, 1 M NaCl, 1.5 mM MgCl₂, 4% glycerol, 0.75 mM DTT; for hydrophobic interaction chromatography HIC'A': 10 mM Tris pH 7.5, 1 M NaCl, 1.5 mM MgCl₂, 4% glycerol, 0.75 mM DTT; HIC'B': 10 mM Tris pH 7.5, 10 mM NaCl, 1.5 mM MgCl₂, 4% glycerol, 0.75 mM DTT. For size chromatography, buffer IEX'A' was used.

C. elegans lacZ reporter assays

pkIs2084 was generated by X-ray mediated integration of the extrachromosomal transgene of strain CT5²³. Note that *pkIs2084* often drives lacZ expression in a few (4–5) seam cells in the head region in a wild-type background. Adult animals are scored positive for a defect in *let-7*-mediated silencing only if the staining extends towards the tail. As control animals, either wild-type or *rrf-3(pk1426)* mutant animals were used. Animals injected with dsRNA against a gene not related to miRNAs (*mut-7*) were also used as a control, with identical results. Animals carrying *pkIs2084* were crossed with animals carrying *drsh-1(tm0654)*, *dcr-1(s2795)* or *pash-1(pk2083)*. Progeny of injected animals and progeny of animals homozygous for the transgene and heterozygous for *tm0654*, *pk2083* or *s2795* were fixed and stained with X-gal. The genotype of stained animals was verified by PCR followed by restriction digest or sequencing. *pash-1(pk2083)* was obtained by target-selected mutagenesis in *C. elegans* (E. Cuppen, unpublished results), and introduces a *Psi* site. *pk2083* changes the TTA codon for L527 into TAA, introducing a premature stop.

Received 9 August; accepted 16 September 2004; doi:10.1038/nature03049.

Published online 7 November 2004.

- Lee, Y., Jeon, K., Lee, J. T., Kim, S. & Kim, V. N. MicroRNA maturation: stepwise processing and subcellular localization. *EMBO J.* **21**, 4663–4670 (2002).
- Lee, Y. *et al.* The nuclear RNase III Drosha initiates microRNA processing. *Nature* **425**, 415–419 (2003).
- Carrington, J. C. & Ambros, V. Role of microRNAs in plant and animal development. *Science* **301**, 336–338 (2003).
- Yi, R., Qin, Y., Macara, I. G. & Cullen, B. R. Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs. *Genes Dev.* **17**, 3011–3016 (2003).
- Lund, E., Guttlinger, S., Calado, A., Dahlberg, J. E. & Kutay, U. Nuclear export of microRNA precursors. *Science* **303**, 95–98 (2004).
- Bohnsack, M. T., Czapinski, K. & Gorlich, D. Exportin-5 is a RanGTP-dependent dsRNA-binding protein that mediates nuclear export of pre-miRNAs. *RNA* **10**, 185–191 (2004).
- Hutvagner, G. *et al.* A cellular function for the RNA-interference enzyme Dicer in the maturation of the *let-7* small temporal RNA. *Science* **293**, 834–838 (2001).
- Grishok, A. *et al.* Genes and mechanisms related to RNA interference regulate expression of the small temporal RNAs that control *C. elegans* developmental timing. *Cell* **106**, 23–34 (2001).
- Ketting, R. F. *et al.* Dicer functions in RNA interference and in synthesis of small RNA involved in developmental timing in *C. elegans*. *Genes Dev.* **15**, 2654–2659 (2001).
- Song, J. J. *et al.* The crystal structure of the Argonaute2 PAZ domain reveals an RNA binding motif in RNAi effector complexes. *Nature Struct. Biol.* **10**, 1026–1032 (2003).
- Zhang, H., Kolb, F. A., Jaskiewicz, L., Westhof, E. & Filipowicz, W. Single processing center models for human Dicer and bacterial RNase III. *Cell* **118**, 57–68 (2004).

12. Siolas, D. *et al.* Synthetic shRNAs as highly potent RNAi triggers. *Nature Biotechnol.* (submitted).
13. Schwarz, D. S. *et al.* Asymmetry in the assembly of the RNAi enzyme complex. *Cell* **115**, 199–208 (2003).
14. Llave, C., Xie, Z., Kasschau, K. D. & Carrington, J. C. Cleavage of Scarecrow-like mRNA targets directed by a class of *Arabidopsis* miRNA. *Science* **297**, 2053–2056 (2002).
15. Yekta, S., Shih, I. H. & Bartel, D. P. MicroRNA-directed cleavage of *HOXB8* mRNA. *Science* **304**, 594–596 (2004).
16. Olsen, P. H. & Ambros, V. The *lin-4* regulatory RNA controls developmental timing in *Caenorhabditis elegans* by blocking LIN-14 protein synthesis after the initiation of translation. *Dev. Biol.* **216**, 671–680 (1999).
17. Zeng, Y., Wagner, E. J. & Cullen, B. R. Both natural and designed micro RNAs can inhibit the expression of cognate mRNAs when expressed in human cells. *Mol. Cell* **9**, 1327–1333 (2002).
18. Doench, J. G., Petersen, C. P. & Sharp, P. A. siRNAs can function as miRNAs. *Genes Dev.* **17**, 438–442 (2003).
19. Wu, H., Xu, H., Miraglia, L. J. & Crooke, S. T. Human RNase III is a 160-kDa protein involved in preribosomal RNA processing. *J. Biol. Chem.* **275**, 36957–36965 (2000).
20. Bernstein, E., Caudy, A. A., Hammond, S. M. & Hannon, G. J. Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature* **409**, 363–366 (2001).
21. Giot, L. *et al.* A protein interaction map of *Drosophila melanogaster*. *Science* **302**, 1727–1736 (2003).
22. Shiohama, A., Sasaki, T., Noda, S., Minoshima, S. & Shimizu, N. Molecular cloning and expression analysis of a novel gene *DGCR8* located in the DiGeorge syndrome chromosomal region. *Biochem. Biophys. Res. Commun.* **304**, 184–190 (2003).
23. Reinhart, B. J. *et al.* The 21-nucleotide *let-7* RNA regulates developmental timing in *Caenorhabditis elegans*. *Nature* **403**, 901–906 (2000).
24. Caudy, A. A. *et al.* A micrococcal nuclease homologue in RNAi effector complexes. *Nature* **425**, 411–414 (2003).
25. Gregory, R. I. *et al.* The Microprocessor complex mediates the genesis of microRNAs. *Nature* doi:10.1038/nature03120 (this issue).
26. Han, M. H., Goud, S., Song, L. & Fedoroff, N. The *Arabidopsis* double-stranded RNA-binding protein HYL1 plays a role in microRNA-mediated gene regulation. *Proc. Natl Acad. Sci. USA* **101**, 1093–1098 (2004).
27. Liu, Q. *et al.* R2D2, a bridge between the initiation and effector steps of the *Drosophila* RNAi pathway. *Science* **301**, 1921–1925 (2003).
28. Tabara, H., Yigit, E., Siomi, H. & Mello, C. C. The dsRNA binding protein RDE-4 interacts with RDE-1, DCR-1, and a DExH-box helicase to direct RNAi in *C. elegans*. *Cell* **109**, 861–871 (2002).
29. Caudy, A. A., Myers, M., Hannon, G. J. & Hammond, S. M. Fragile X-related protein and VIG associate with the RNA interference machinery. *Genes Dev.* **16**, 2491–2496 (2002).
30. Spector, D. L. & Smith, H. C. Redistribution of U-snRNPs during mitosis. *Exp. Cell Res.* **163**, 87–94 (1986).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank F. Rivas for critical reading of the manuscript. Strain BC3825 was obtained from the *C. elegans* Genetics Center, and *drsh-1(tm0654)* was obtained from the NBP in Japan (Mitani laboratory). We thank E. Cuppen and his group for help in target-selected mutagenesis. A.M.D. is a David Koch Fellow of the Watson School of Biological Sciences. G.J.H. is supported by an Innovator Award from the US Army Breast Cancer Research Program. This work was also supported by a grant from the NIH (G.J.H.) and by a VENI fellowship from the Netherlands Organisation for Scientific Research (R.E.K.).

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to G.H. (hannon@cshl.org).

The Microprocessor complex mediates the genesis of microRNAs

Richard I. Gregory*, Kai-ping Yan*, Govindasamy Amuthan, Thimmaiah Chendrimada, Behzad Doratotaj, Neil Cooch & Ramin Shiekhattar

The Wistar Institute, 3601 Spruce Street, Philadelphia, Pennsylvania 19104, USA

* These authors contributed equally to this work

MicroRNAs (miRNAs) are a growing family of small non-protein-coding regulatory genes that regulate the expression of homologous target-gene transcripts. They have been implicated in the control of cell death and proliferation in flies^{1,2}, haematopoietic lineage differentiation in mammals³, neuronal patterning in nematodes⁴ and leaf and flower development in plants^{5–8}. miRNAs are processed by the RNA-mediated interference machinery. Drosha is an RNase III enzyme that was recently

implicated in miRNA processing. Here we show that human Drosha is a component of two multi-protein complexes. The larger complex contains multiple classes of RNA-associated proteins including RNA helicases, proteins that bind double-stranded RNA, novel heterogeneous nuclear ribonucleoproteins and the Ewing's sarcoma family of proteins. The smaller complex is composed of Drosha and the double-stranded-RNA-binding protein, DGCR8, the product of a gene deleted in DiGeorge syndrome. *In vivo* knock-down and *in vitro* reconstitution studies revealed that both components of this smaller complex, termed Microprocessor, are necessary and sufficient in mediating the genesis of miRNAs from the primary miRNA transcript.

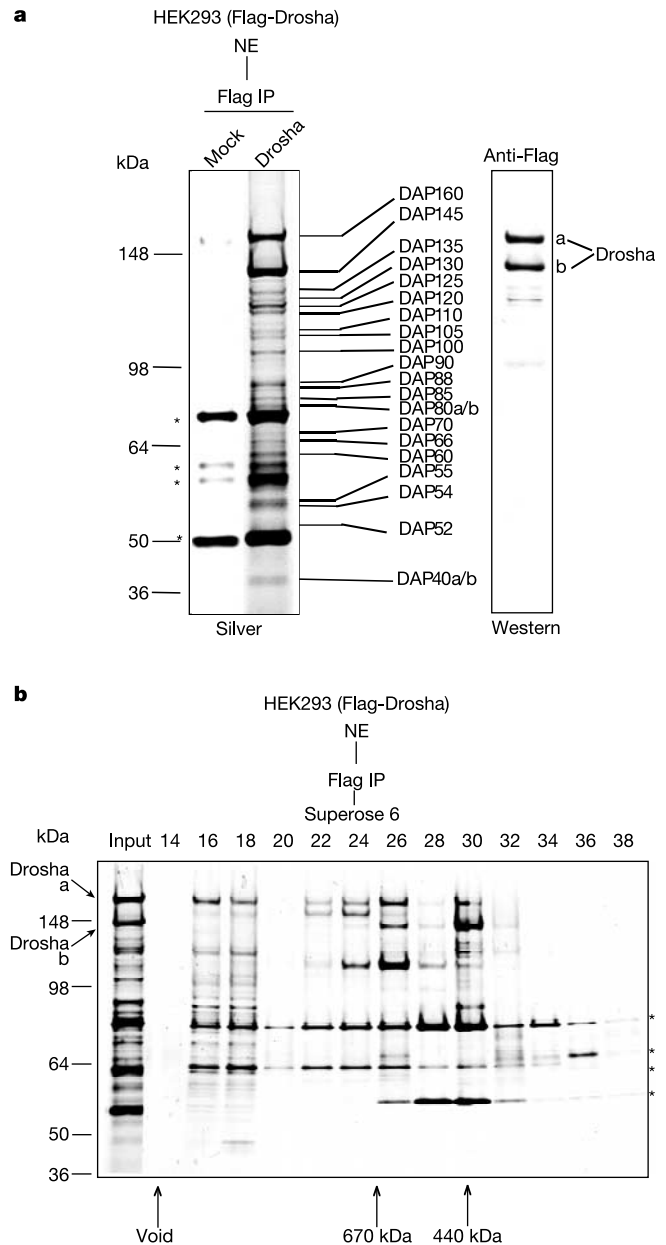


Figure 1 Isolation of Drosha-containing complexes. **a**, Fractions of the immunoaffinity eluate from M2 anti-Flag beads resolved by SDS-PAGE (4–12%). Flag-Drosha was revealed by silver staining and western blotting with anti-Flag antibodies. Molecular masses of marker proteins (left) and the polypeptide masses of associated subunits (right) are indicated. **b**, Silver staining of Superose 6 gel-filtration fractions. Top, fraction number; bottom, molecular mass markers. DAP, Drosha-associated proteins having different molecular masses; asterisks, contaminating polypeptides; IP, immunoprecipitation.

It was recently suggested that most miRNA genes originate from independent transcription units^{9–11}. However, about a quarter of human miRNA genes are located in introns of pre-mRNAs. Because these miRNAs have the same orientation as pre-mRNAs, it is likely that they are not transcribed from their own promoters but are processed from the introns^{12–14}. The remaining miRNAs are clustered in the genome, predicting a multi-cistronic transcript^{9,10}. Regardless of how different miRNAs originate, the primary miRNA transcript (pri-miRNA)¹⁵ must be processed to yield a mature 22-nucleotide (nt) miRNA. The present model for the pathway by which mammalian miRNAs undergo maturation begins with cleavage of the pri-miRNA in the nucleus to release a ~60–70-nt stem loop intermediate, known as the pre-miRNA^{15,16}. This processing event is mediated by an RNase III endonuclease, Drosha, which cleaves both strands of the stem at sites near the base of the primary stem loop¹⁷. Whether Drosha in itself is sufficient for this processing or whether other components associated with Drosha contribute to the cleavage of pri-miRNAs has not previously been addressed.

To gain insight into the components of the miRNA-processing machinery, we isolated a Drosha-containing complex from human cells. This was accomplished by developing HEK-293-derived stable cell lines expressing Flag-tagged Drosha. Flag-Drosha was isolated by immunoaffinity chromatography. As Fig. 1a demonstrates,

the Flag affinity eluate contains a rich harvest of polypeptides. Drosha was represented in two forms differing in molecular mass by ~10–15 kDa (Drosha a, ~160 kDa, and Drosha b, ~145 kDa; see Fig. 1a). It is currently not clear whether these two forms reflect post-translational modification of Drosha or result from proteolytic cleavage sustained during preparation of the nuclear extract.

To characterize the elution profiles of the two forms of Drosha and to demonstrate that the Drosha-associated polypeptides (DAPs) constitute a multi-protein complex, the Flag-affinity eluate was fractionated on a gel-filtration column (Fig. 1b). This analysis revealed the presence of the larger form of Drosha (Drosha a) in two distinct complexes: a high-molecular-mass complex containing most Drosha-associated polypeptides (fractions 16–18) and a lower-molecular-mass complex (fractions 22–26, ~600 kDa). The smaller form of Drosha (Drosha b) displayed an elution profile consistent with a smaller complex with a molecular mass of ~400 kDa (fractions 30–32).

We next determined the identity of the Drosha-associated polypeptides. The Flag-affinity eluate was separated on a polyacrylamide gel, bands were stained with colloidal blue, and individual polypeptides were excised from the gel and subjected to mass spectrometric sequencing. Nineteen specific Drosha-associated polypeptides were identified in two independent sequencing

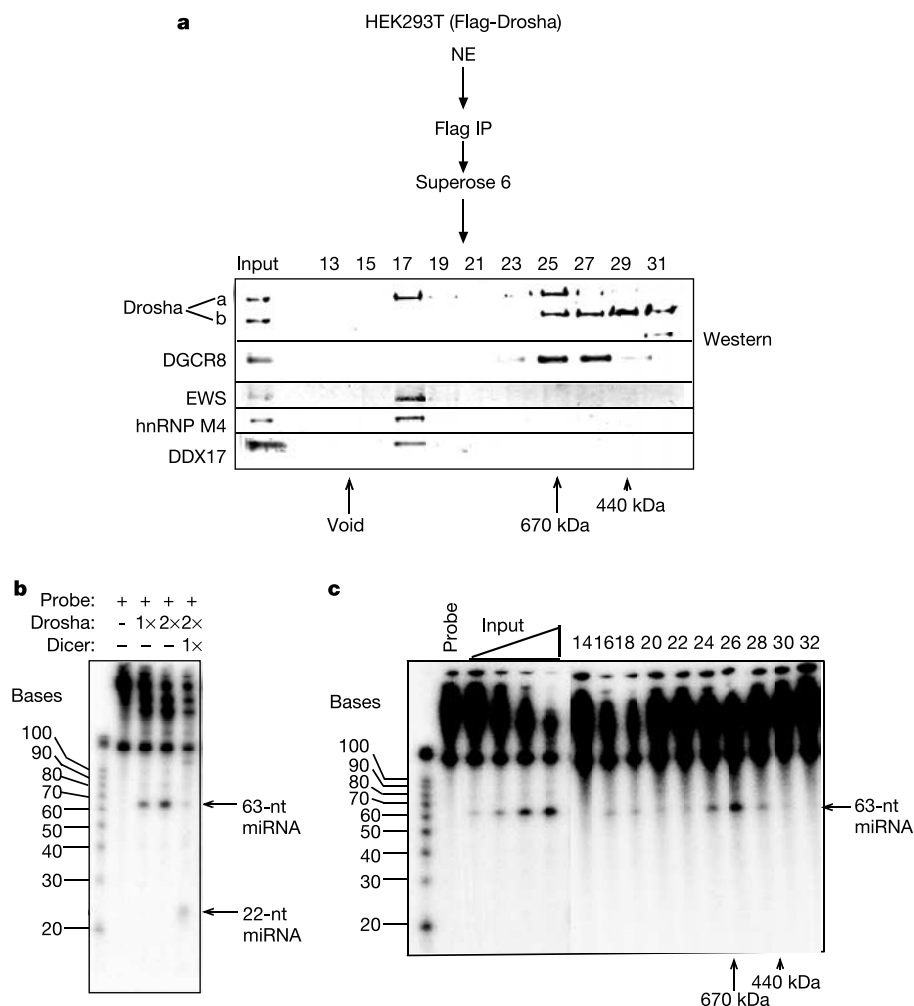


Figure 2 Composition and miRNA processing activity of Superose 6 gel-filtration fractions. **a**, Western blotting with the antibodies shown at the left. IP, immunoprecipitation. **b**, Flag-affinity eluates and an enriched Dicer fraction were used for miRNA processing. '1 ×' corresponds to 10 ng of Drosha and 5 μ l of enriched Dicer

fraction. **c**, miRNA processing with a pri-miRNA miR-(17,18,19a,20,19b-1) fragment. Flag-Drosha eluate (Input) corresponded to 5, 10, 20 and 40 ng of Drosha, determined as described in Methods.

analyses (Supplementary Fig. S1). The Drosha-associated proteins comprised specific classes of RNA-associated proteins displaying common structural domains. These included the DEAD-box and DEAH-box family of RNA helicases, proteins with domains that bind double-stranded RNA, heterogeneous nuclear ribonucleoproteins (hnRNPs) and the Ewing's sarcoma family of proteins containing an RNA recognition motif (RRM) and a zinc-finger domain.

To address the association of the aforementioned different classes of polypeptides with Drosha, we analysed alternate fractions of Superose 6 gel-filtration chromatography by western blot analysis. This analysis revealed that Drosha coelutes with the protein product of Ewing's sarcoma gene (EWS), the RNA helicase, DDX17/P72 and hnRNPM4 in a large complex peaking in fraction 17 (Fig. 2a). In contrast with these polypeptides, a single polypeptide of 120 kDa corresponding to DGCR8 coelutes with the smaller

complex peaking in fractions 25–27 (Figs 1b and 2a).

We next analysed the Flag-affinity eluate for pri-miRNA processing. A 790-base-pair fragment corresponding to a cluster of pri-miRNAs miR-(17,18,19a,20,19b-1) was used as a substrate for analysis of pri-miRNA processing. It was shown recently that pri-miRNA processing catalysed by Drosha resulted in the formation of a ~60–70-base-pair pre-miRNA precursor¹⁷. We first examined the activity of the Flag-Drosha affinity eluate in processing the pri-miRNA. Addition of increasing concentrations of Flag-Drosha affinity eluate to the pri-miRNA processing reaction resulted in the appearance of a distinct 63-nt pre-miRNA fragment (Fig. 2b). Further addition of a fraction enriched for Dicer, which catalyses further processing of the pre-miRNA, converted this fragment to a mature ~22-nt miRNA, confirming that this pre-miRNA is the correct processing product (Fig. 2b).

We next analysed the gel-filtration fractions for pri-miRNA processing activity and found it was present in two distinct peaks corresponding to the two Drosha-containing complexes (Fig. 2c). Although the larger complex (fractions 16–18) displayed some pri-miRNA processing activity, the bulk of pri-miRNA processing activity co-eluted with the smaller Drosha complex containing DGCR8. Moreover, closer examination of the reaction revealed that the fractions containing the larger Drosha complex contained a non-specific RNase activity, resulting in reduced probe concentrations.

To characterize the polypeptide composition of the smaller Drosha-containing complex, fractions 24–27 were concentrated by trichloroacetic acid, treated as described above and subjected to mass spectrometric sequencing. The three largest polypeptides corresponded to Drosha a and b forms and the DGCR8 protein (Supplementary Fig. S2). (The smaller polypeptide corresponded to SKB1, a common contaminant of Flag-affinity purification.) We then analysed the specific pri-miRNA processing activity of the two Drosha complexes by normalizing the amounts of Drosha in each complex (Supplementary Fig. S2). These studies revealed that the Drosha–DGCR8 complex displays a nearly eightfold greater pri-miRNA processing activity than the large Drosha complex (Supplementary Fig. S2).

We next developed HEK-293-derived stable cell lines expressing Flag-DGCR8. Flag-DGCR8 was isolated by affinity chromatography. As shown in Fig. 3a, the DGCR8 affinity eluate contained a single polypeptide corresponding to Drosha in addition to DGCR8 itself. The polypeptide below DGCR8 corresponds to a carboxy-terminal truncation of the protein, because only the amino-terminal DGCR8 antibodies recognize this truncated protein (Fig. 3a). The Flag-DGCR8–Drosha complex was then used to assess its activity for pri-miRNA processing. Consistent with the results obtained with the smaller Drosha complex (Supplementary Fig. S2), Flag-DGCR8–Drosha exhibited robust pri-miRNA processing activity (Fig. 3b). These results demonstrate the stable association of Drosha and DGCR8 in an active pri-miRNA processing complex. For the sake of consistency we have called this complex Microprocessor¹⁸. A parallel study shows that a *Drosophila* homologue of DGCR8 associates with Drosha and pri-miRNA processing activity in *Drosophila*, *Caenorhabditis elegans* and mammals for a role of this protein in miRNA biogenesis¹⁸.

We next analysed the activity of Microprocessor purified by either Flag-Drosha or Flag-DGCR8 affinity chromatography for processing of two other pri-miRNAs. miR-(15,16) and miR-(23,27,24-2) were processed to yield expected miRNA precursors as was shown previously¹⁵ (Fig. 3c). Moreover, analysis of the processed fragments by northern blot analysis and further processing by Dicer confirmed the specific processing by Microprocessor (Supplementary Fig. S3).

To demonstrate rigorously that pri-miRNA processing required DGCR8 in addition to the RNase III Drosha, we reconstituted pri-miRNA processing activity by using recombinant Drosha produced in insect cells and recombinant DGCR8 generated in

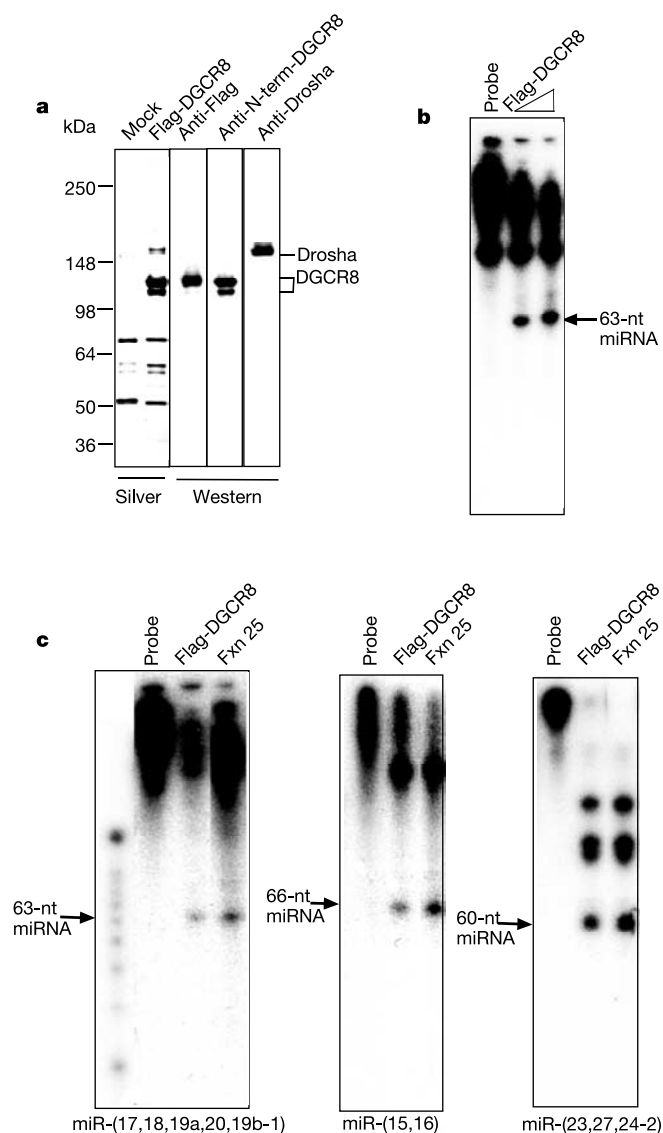


Figure 3 Microprocessor is the Drosha–DGCR8 complex mediating miRNA processing. **a**, Isolation of Flag-DGCR8 complex. SDS-PAGE followed by silver staining and western blot analysis with antibodies shown on the figure are displayed. **b**, Analysis of pri-miRNA processing activity of Flag-DGCR8–Drosha complex. **c**, Analysis of miRNA processing activity of Microprocessor purified through Flag-Drosha (fraction 25; Fxn 25) or Flag-DGCR8 with the use of the miRNAs shown.

bacteria (Fig. 4a). Both recombinant proteins were purified to near homogeneity and were used in near-stoichiometric amounts in pri-miRNA processing assays (Fig. 4b, c). Although native Microprocessor displayed robust pri-miRNA processing activity, neither recombinant Drosha nor recombinant DGCR8 showed any significant pri-miRNA processing activity (Fig. 4c, lanes 4–7). However, addition of the two recombinant proteins reconstituted the pri-miRNA processing activity to levels similar to those seen with native complex (Fig. 4c, lanes 8 and 9). Whereas the addition of increasing concentrations of DGCR8 inhibited the processing reaction perhaps through a squelching mechanism, a further increase in Drosha stimulated the processing of pri-miRNA (Fig. 4c, compare lanes 10, 11 and 12, 13). Interestingly, whereas Drosha alone displayed non-specific RNase activity on the substrate, the addition of DGCR8 inhibited these non-specific effects and promoted Drosha's pri-miRNA processing activity (Fig. 4c, compare lanes 5 and 9). These results show the requirement for DGCR8 in directing the specific processing of pri-miRNAs by Drosha.

To assess the role of Microprocessor in the initiation of miRNA processing *in vivo*, we used RNA interference to deplete DGCR8 and

Drosha. For these experiments short interfering RNA (siRNA) against Drosha was used as a positive control, whereas siRNA against transcription factor TFII-I was used as negative control (Fig. 5a). Knock-down of DGCR8 caused a similar effect to Drosha depletion for all miRNAs tested, resulting in a pronounced decrease in mature miRNA levels (Fig. 5b). Depletion of both Drosha and DGCR8, however, resulted in a substantial accumulation of pri-miRNA, showing the requirement for Microprocessor in miRNA processing *in vivo* (Fig. 5c). These results show the obligatory role for Drosha and DGCR8 in microRNA processing *in vivo* and *in vitro*.

We have isolated two multi-protein complexes that contain Drosha as their catalytic engine. We show that the smaller Microprocessor complex containing Drosha and DGCR8 is necessary and sufficient for the processing of pri-miRNA to pre-miRNA. This contention is based on the following observations. First, Microprocessor purified by Flag-Drosha affinity purification displays specific and robust activity in pri-miRNA processing. Second, isolation of Microprocessor through Flag-DGCR8 using stable cell lines revealed its close association with Drosha and miRNA

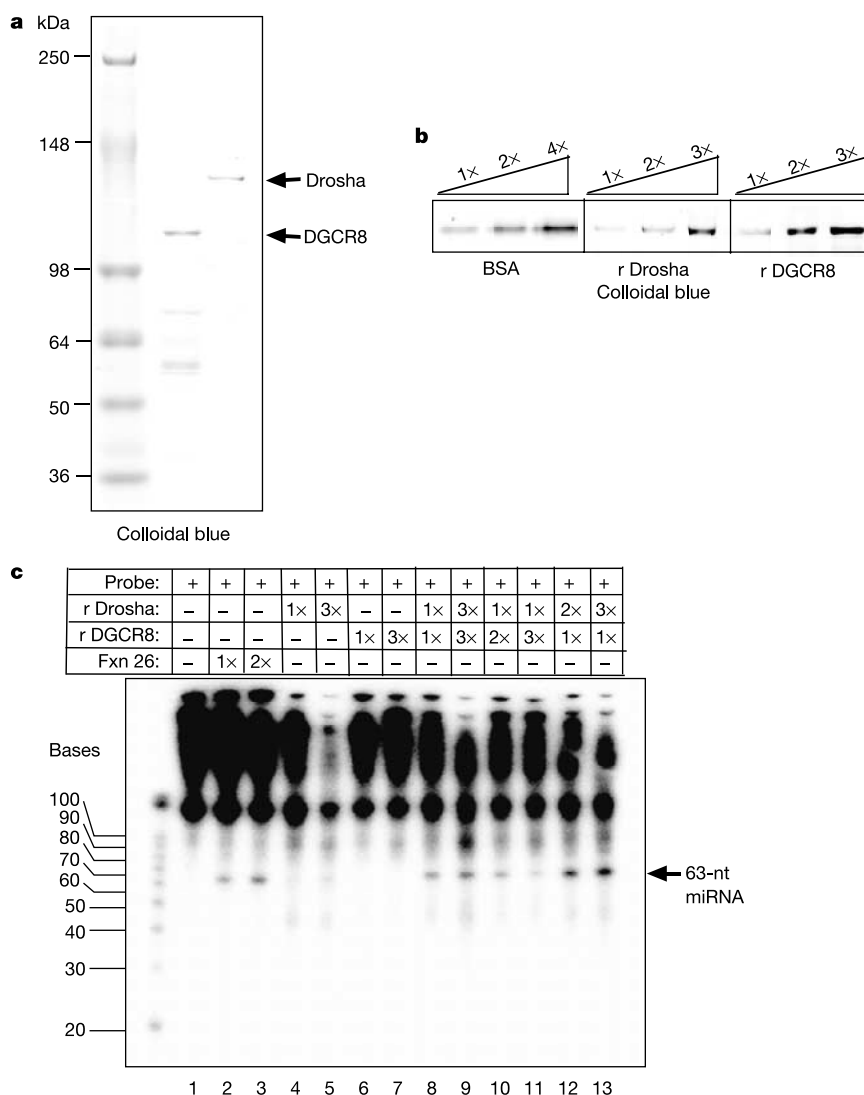


Figure 4 Reconstitution of Microprocessor by using recombinant Drosha and DGCR8. **a**, Analysis of recombinant Drosha and DGCR8 with colloidal blue staining; 100 ng of each protein was analysed. **b**, As in **a** except that '1 ×' corresponds to 50 ng of BSA, used to determine Drosha and DGCR8 concentrations. **c**, Reconstitution of miRNA processing by

using recombinant Drosha and DGCR8. '1 ×' corresponds to 10 ng of Drosha in fraction 26 (Fxn 26) of Superose 6 and 20 ng of each of the recombinant (r) proteins assayed for pri-miRNA processing activity as described in Methods.

processing activity. Third, miRNA processing activity could be reconstituted with recombinant Drosha and DGCR8. Last, knock-down of Drosha and DGCR8 resulted in diminished mature miRNA and accumulation of pri-miRNA *in vivo*.

DGCR8 is an evolutionarily conserved protein that contains two double-stranded RNA-binding domains and a WW domain known to interact with proline-rich peptides. The WW domain of DGCR8 is most probably the interacting surface with the proline-rich N-terminal domain of Drosha. DGCR8 is one of an estimated 30 genes in the chromosomal region 22q11.2 whose heterozygous deletion results in the most common human genetic deletion syndrome, known as DiGeorge syndrome^{19,20}. The clinical manifestations of the disease are highly variable, with 75% of patients displaying congenital heart defects. Other common features include, among many others, characteristic facial appearance, immunodeficiency resulting from thymic hypoplasia, and developmental and behavioural problems. It will be important to explain the role of DGCR8 in the genesis and development of DiGeorge syndrome. Future experiments with knockout mouse models of DGCR8 in homozygous and heterozygous animals will shed light on the likely function of DGCR8 in developmental control and the expression of DiGeorge syndrome-like phenotypes.

In the past we have used the 2-MDa chromatin remodelling complex, SWI-SNF, as a size marker for gel-filtration chromatography. Comparing the elution profiles of the large Drosha complex with that of SWI-SNF (which peaks in fraction 20 on

Superose 6) we conclude that the large Drosha complex is larger than SWI-SNF. We have identified nearly 20 polypeptides that are specifically associated with Drosha in this complex. These included RNA helicases containing a DEAD-box or DEAH-box, hnRNPs and proteins with either double-stranded RNA-binding or RRM domains. We have also identified EWS as components of this larger Drosha complex. Ewing's family of tumours result from tumour-associated chromosomal translocations between EWS genes and one of five different ETS transcription factors²¹.

We cannot exclude a role for the large Drosha complex in miRNA processing because it displayed a weak pre-miRNA processing activity *in vitro*. Moreover, our analysis *in vivo* with siRNA knock-down of three components, p62/DDX5, p72/DDX17 and hnRNP1-like, revealed a small decrease in mature miRNAs, although abrogation of these subunits never approached the effects observed after DGCR8 and Drosha depletions. It is therefore more likely that the large Drosha-containing complex has a function in other RNA processing pathways. Because Drosha has also been previously shown to participate in preribosomal RNA processing²², the large Drosha complex might mediate such preribosomal RNA processing activities. □

Methods

Affinity purification of Flag-Drosha and Flag-DGCR8

Flag-Drosha and a selectable marker for puromycin resistance were cotransfected into HEK-293 human embryonic kidney cells. Transfected cells were grown in the presence of 2.5 $\mu\text{g ml}^{-1}$ puromycin, and individual colonies were isolated and analysed for Flag-Drosha expression. To purify the Drosha complex, nuclear extract generated from 200 15-cm plates (4×10^9 cells or ~ 150 mg of nuclear extract) was incubated with anti-Flag M2 affinity resin (Sigma). After two washes with buffer A (20 mM Tris-HCl pH 7.9, 0.5 M KCl, 10% glycerol, 1 mM EDTA, 5 mM dithiothreitol, 0.5% Nonidet P40, 0.2 mM phenylmethylsulphonyl fluoride), the affinity column was eluted with buffer A containing Flag peptide (400 $\mu\text{g ml}^{-1}$) in accordance with the manufacturer's instructions (Sigma). Analysis of Drosha by Superose 6 gel filtration was similar to that previously described^{23,24}. Fractions from the gel-filtration chromatography were concentrated by precipitation with trichloroacetic acid and analysed by SDS-PAGE followed by silver staining. Flag-DGCR8 was isolated by using a similar protocol. Protein identification by liquid chromatography-tandem mass spectrometry was performed as detailed²³. Multiple sequencing analyses have determined SKB1, MEP50 and α -tubulin as common contaminants of Flag-affinity purification. Anti-hnRNP M4 antibody was obtained from Santa Cruz Biotechnology. Anti-Drosha and anti-DGCR8 were generated against the last 20 amino acids in the amino and carboxy termini of each protein (Open Biosystems). Recombinant Drosha and DGCR8 were purified by methods previously described for the purification of recombinant proteins from insect cells and bacteria^{23,24}. In brief, each protein was expressed as a Flag-tagged protein and purified with anti-Flag M2 affinity resin similar in manner to the protocol described for Flag-Drosha and Flag-DGCR8.

miRNA processing

In vitro transcription was performed with the Promega Riboprobe system, using linearized pGEM-7Z vector containing miR-(17,18,19a,20,19b-1), miR-(15,16) and miR-(23,27,24-2) as described¹⁵. In brief, the miRNA probes were amplified from HEK-293 RNA by reverse transcriptase-mediated polymerase chain reaction (RT-PCR) with 5'-tgctgaattgtatggtttatagttgta-3' as 5' primer and 5'-tactttctacagacttttactaccaca-3' as 3' primer for miR-(17,18,19a,20,19b-1), 5'-CGCCCGGTGCCCCCTCACCCCTGTGC CAC-3' as 5' primer and 5'-CCCTGTTCTGCTGAAGTGAAGCCAGTGTAC-3' as 3' primer for miR-(23,27,24-2) and 5'-CCTTGAGTAAAGTAGCAGCAACTAATG-3' as 5' primer and 5'-CTTACTCTGAGTTAAATCTTGAATAC-3' as 3' primer for miR-(15,16). The processing reaction, containing the indicated amounts of Drosha complex, 3 μl of solution containing 32 mM MgCl_2 , 10 mM ATP, 200 mM creatine phosphate, 1 U μl^{-1} HPRI (Takara) and the labelled pri-miRNA (2×10^5 c.p.m.) and buffer (20 mM Tris-HCl pH 7.9, 0.1 M KCl, 10% glycerol, 5 mM dithiothreitol, 0.2 mM phenylmethylsulphonyl fluoride), was added to a final volume of 30 μl . The reaction mixture was incubated at 37°C for 90 min and extracted with phenol:chloroform mixture, then with chloroform and precipitated with 300 mM sodium acetate and ethanol. The precipitated RNA was loaded on 15% denaturing polyacrylamide gels. For the reconstitution experiments, the two recombinant proteins were incubated for 1 h on ice in buffer A containing 100 mM KCl before the addition of the reaction mix. All Drosha quantitative analyses were performed by comparing the Flag-affinity eluate and recombinant Drosha by quantitative western blot analysis. The amounts of recombinant Drosha were then deduced by using colloidal blue staining in comparison with known amounts of BSA.

siRNAs and transfection

The siRNAs were synthesized by Dharmacon. The sequence of Drosha siRNA was 5'-AAGGACCAAGUUAUCAGCAAG-3', the DGCR8 siRNA was 5'-AUCCGUUGAUCUC GAGGAATT-3', and the control siRNA against TFII was 5'-UGUGGGGAAGCUCUU

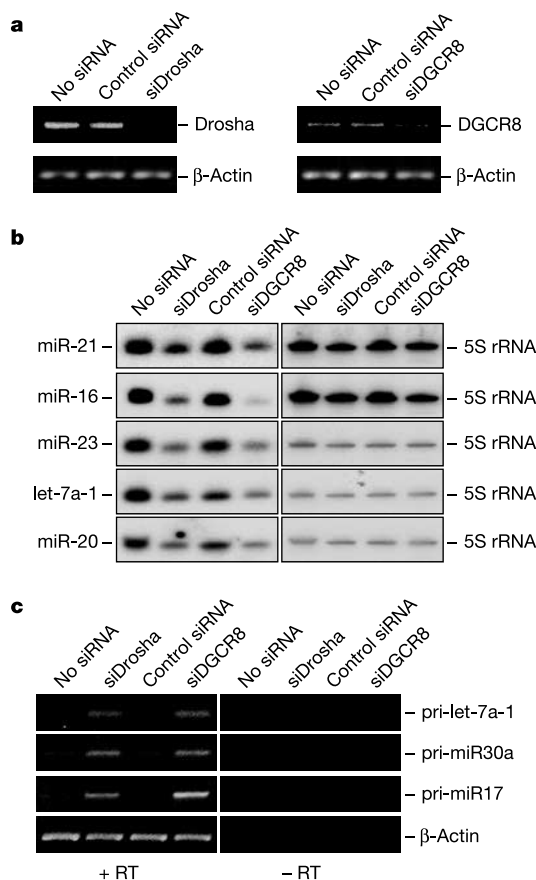


Figure 5 A role for Microprocessor *in vivo* in miRNA processing. **a**, Analysis of transcript levels by using RT-PCR for Drosha and DGCR8 after treatment of HeLa cells with siRNA against each protein. siRNA against TFII-1 was used as control. **b**, Northern blot analysis of miRNA-21, miRNA-16, miRNA-23, let-7a-1 and miRNA-20 after treatment of HeLa cells with siRNA against Drosha, DGCR8 or control siRNA. **c**, Analysis of pri-miRNA processing after depletion of Drosha and DGCR8. —RT, absence of reverse transcriptase.

GGCCTT-3'. siRNA transfection in HeLa cells was performed with Lipofectamine 2000 (Invitrogen). In brief, cells were plated in 10-cm dishes to 40% confluence. For each dish, 1.6 nmol of siRNA was mixed with 24 µl of Lipofectamine 2000 in 3 ml of Opti-MEM medium. The mixture was added to cells and incubated for 6 h. After 24 h a second transfection was performed in the same way. Total RNA was prepared 3 days after the second transfection and was used for northern blot analysis or RT-PCR.

RNA isolation, RT-PCR and northern blot analysis

Total RNA from HeLa cells was prepared in TRIzol reagents (Invitrogen) in accordance with the manufacturer's instructions. To examine the effect of siRNA, RNA (2 µg) was subjected to complementary DNA synthesis with oligo(dT), using the SuperScript first-strand synthesis system for RT-PCR (Invitrogen). To examine pri-miRNAs, RNA (2 µg) was subjected to cDNA synthesis with random primers. As a control, RT-PCRs were performed in the absence of reverse transcriptase. Different PCR cycles were examined to determine linear amplification.

Primer sequences for RT-PCRs in Fig. 5a were *Drosha* (5'-CATGCACCAGATTCTCCT GTA-3' and 5'-GTCTCCTGCATAACTCAACTG-3') and *DGCR8* (5'-TATCAGATCC TCCACGAGTG-3' and 5'-TCTTGGAGCTTGCTGAGGAT-3'). Primer sequences for RT-PCRs in Fig. 5c were pri-let-7a-1 (5'-GATTCTTTTCACCATTCACCTGGA TGTT-3' and 5'-TTTCTATCAGACCGCTGGATGCAGACTTT-3'), pri-miR30a (5'-ATTGCTGTTTGAATGAGGCTTCAGTACTTT-3' and 5'-TTCAGCTTTGTAAAAA TGTATCAAAGAGAT-3') and pri-miR17 (5'-TGCTGAATTGTATGGTTTATAGTTG TTAG-3' and 5'-CACTACCACAGTCAGTTTGCATGGATTG-3'). β-Actin was used for internal control in Fig. 5a, c, and the primer sequences were 5'-AAAGACCTGTA CGCCAAACAC-3' and 5'-GTCATACTCCTGCTTGCTGAT-3'.

For northern blot analysis, total RNA (10 µg) was resolved on a 15% denaturing polyacrylamide gel and electrotransferred to Hybond N⁺ nylon membrane (Amersham). The membranes were crosslinked under ultraviolet and prewashed for 1 h at 65 °C in 0.1 × SSC/0.1% SDS. Prehybridization and hybridization were performed at 42 °C in 10 × Denharts solution, 6 × SSC, 0.1% SDS. Oligonucleotides complementary to miRNAs were end-labelled with [γ -³²P]ATP and used as probes for northern analysis. The sequences of the oligonucleotides were 5'-gaaatccctggcaatgtgat-3' (miR-23), 5'-actatacaactactactca-3' (let-7a-1), 5'-gccaatattactgtctgcta-3' (miR-16), 5'-tacctgcactataagcacttta-3' (miR-20), 5'-tcaacatcagctctgataagcta-3' (miR-21) and 5'-caggcccgaccctgcttagcttcagatcagacagat-3' (5S rRNA). All of the probes were washed twice for 10 min at 25 °C in 6 × SSC/0.1% SDS.

Received 14 June; accepted 19 October 2004; doi:10.1038/nature03120.

Published online 7 November 2004.

1. Brennecke, J. *et al.* *bantam* encodes a developmentally regulated microRNA that controls cell proliferation and regulates the proapoptotic gene *hid* in *Drosophila*. *Cell* **113**, 25–36 (2003).
2. Xu, P. *et al.* The *Drosophila* microRNA *mir-14* suppresses cell death and is required for normal fat metabolism. *Curr. Biol.* **13**, 790–795 (2003).
3. Chen, C. Z. *et al.* MicroRNAs modulate hematopoietic lineage differentiation. *Science* **303**, 83–86 (2004).
4. Johnston, R. J. & Hobert, O. A microRNA controlling left/right neuronal asymmetry in *Caenorhabditis elegans*. *Nature* **426**, 845–849 (2003).
5. Aukerman, M. J. & Sakai, H. Regulation of flowering time and floral organ identity by a microRNA and its *APETALA2*-like target genes. *Plant Cell* **15**, 2730–2741 (2003).

6. Chen, X. *et al.* A microRNA as a translational repressor of *APETALA2* in *Arabidopsis* flower development. *Science* **303**, 2022–2025 (2004).
7. Emery, J. F. *et al.* Radial patterning of *Arabidopsis* shoots by class III *HD-ZIP* and *KANADI* genes. *Curr. Biol.* **13**, 1768–1774 (2003).
8. Palatnik, J. F. *et al.* Control of leaf morphogenesis by microRNAs. *Nature* **425**, 257–263 (2003).
9. Lagos-Quintana, M. *et al.* Identification of novel genes coding for small expressed RNAs. *Science* **294**, 853–858 (2001).
10. Lau, N. C. *et al.* An abundant class of tiny RNAs with probable regulatory roles in *Caenorhabditis elegans*. *Science* **294**, 858–862 (2001).
11. Lee, R. C. & Ambros, V. An extensive class of small RNAs in *Caenorhabditis elegans*. *Science* **294**, 862–864 (2001).
12. Aravin, A. A. *et al.* The small RNA profile during *Drosophila melanogaster* development. *Dev. Cell* **5**, 337–350 (2003).
13. Lagos-Quintana, M. *et al.* New microRNAs from mouse and human. *RNA* **9**, 175–179 (2003).
14. Lim, L. P. *et al.* The microRNAs of *Caenorhabditis elegans*. *Genes Dev.* **17**, 991–1008 (2003).
15. Lee, Y., Jeon, K., Lee, J. T., Kim, S. & Kim, V. N. MicroRNA maturation: stepwise processing and subcellular localization. *EMBO J.* **21**, 4663–4670 (2002).
16. Zeng, Y. & Cullen, B. R. MicroRNAs and small interfering RNAs can inhibit mRNA expression by similar mechanisms. *Proc. Natl Acad. Sci. USA* **100**, 9779–9784 (2003).
17. Lee, Y. *et al.* The nuclear RNase III Drosha initiates microRNA processing. *Nature* **425**, 415–419 (2003).
18. Denali, A. M., Tops, B. B. J., Plasterk, R. H. A., Ketting, R. F. & Hannon, G. J. Processing of primary microRNAs by the Microprocessor complex. *Nature* doi:10.1038/nature03049 (this issue).
19. Yamagishi, H. & Srivastava, D. Unraveling the genetic and developmental mysteries of 22q11 deletion syndrome. *Trends Mol. Med.* **9**, 383–389 (2003).
20. Shiohama, A., Sasaki, T., Noda, S., Minoshima, S. & Shimizu, N. Molecular cloning and expression analysis of a novel gene *DGCR8* located in DiGeorge syndrome chromosomal region. *Biochem. Biophys. Res. Commun.* **304**, 184–190 (2003).
21. Arvand, A. & Denny, C. T. Biology of EWS/ETS fusions in Ewing's family tumors. *Oncogene* **20**, 5747–5754 (2001).
22. Wu, H., Xu, H., Miraglia, L. J. & Crooke, S. T. Human RNase III is a 160-kDa protein involved in preribosomal RNA processing. *J. Biol. Chem.* **275**, 36957–36965 (2000).
23. Bochar, D. A. *et al.* BRCA1 is associated with a human SWI/SNF-related complex: linking chromatin remodeling to breast cancer. *Cell* **102**, 257–265 (2000).
24. Dong, Y. *et al.* Regulation of BRCC, a holoenzyme complex containing BRCA1 and BRCA2, by a signalosome-like subunit and its role in DNA repair. *Mol. Cell* **12**, 1087–1099 (2003).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank V. N. Kim for the Drosha cDNA; O. Delattre and A. I. Lamond for the gift of EWS and DDX17 antibodies, respectively; K. Nishikura for providing recombinant Drosha; and T. Beer (Wistar Proteomics Facility) for expertise in the microcapillary HPLC/mass spectrometry. R.S. was supported by grants from the NIH and the American Cancer Institute. R.G. is a fellow of the Jane Coffin Child Memorial Fund for Medical Research.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to R.S. (shiekhattar@wistar.upenn.edu).

Making connections

Technological innovations in detecting and studying protein–protein interactions are providing new ways of doing research in cell signalling. Diane Gershon investigates.

Cell-signalling research aims to understand how cells convert extracellular signals into the required cellular responses, and how these pathways can go awry and cause disease. With the emergence of cell signalling as a significant field in its own right over the past few years, a dizzying array of reagents, assays and technologies are now available to help researchers (see ‘In the market place’).

Cell-surface receptors

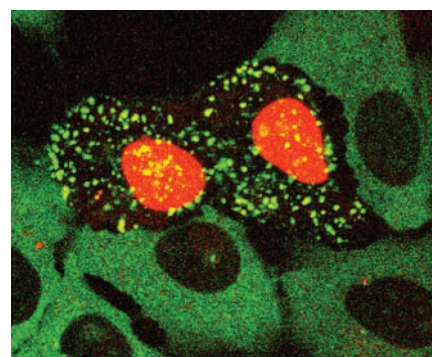
The large and diverse family of G-protein-coupled receptors (GPCRs) responds to a wide range of agonists, including amines, hormones, neurotransmitters and even light. This sheer diversity makes GPCRs a rich source of potential drug targets.

“We spent a lot of the 1990s, as did others, isolating novel genes encoding these GPCRs, which essentially had eluded classical pharmacology,” says Brian O’Dowd of the department of pharmacology at the University of Toronto, Canada. But many of these are still ‘orphan’ GPCRs, for which a natural ligand has not been identified.

To try to tap their therapeutic potential, O’Dowd’s group has developed a multi-purpose cellular assay (MOCA) that can be used to screen for compounds that activate or block any GPCR, and that needs no prior knowledge about the G proteins or second messengers involved. O’Dowd’s lab jokingly refers to this as the ‘mother of all cellular assays’. The technology is being commercialized by Toronto-based PatoBios, of which O’Dowd is a co-founder.

The GPCR is genetically modified to incorporate a nuclear-localization sequence, which causes the receptor to be internalized and translocated to the nucleus, from which it is unable to recycle to the cell surface. The interaction of a ligand with the modified GPCR prevents translocation, and the receptor is retained on the cell surface. The different distributions of the receptor can be visualized by various methods such as fluorescent tags. The assay can also be adapted to detect ligand binding to the important family of transporter proteins involved in the re-uptake of neurotransmitters.

Norak Biosciences of Research Triangle



R.H. OAKLEY/NORAK BIOSCIENCES

Activated receptors show up as dots.

Park, North Carolina, has also developed a cell-based fluorescence assay for finding ligands for orphan GPCRs. Its Transfluor technology detects the binding of β -arrestin to the cytoplasmic part of the receptor, which occurs only after a ligand has bound. Activated receptors can then be detected and isolated. The assay has been validated on various image analysers, including the IN Cell Analyzer 3000 from GE Healthcare, Little Chalfont, UK, the ArrayScan from Cellomics, Pitts-

IN THE MARKET-PLACE

Finding the right reagent or assay for the job in hand can be a daunting prospect in a field as complex as cell signalling. Some companies are taking a one-stop shopping approach; others are pursuing niche markets, specializing in, for example, antibodies, peptide substrates for kinases and proteases, enzymes, especially protein kinases, or protein and peptide microarrays, such as those from Sigma-Aldrich in St Louis, Missouri, PepScan Systems in Lelystad, the Netherlands, and Zeptosens in Witterswil,

Switzerland. Even companies usually thought of as having a molecular biology focus are getting in on the act, as evidenced by the acquisition of Molecular Probes in Eugene, Oregon, a leader in novel fluorescence-based technologies for labelling biological molecules, by Invitrogen of Carlsbad, California in 2003.

“I think it is fair to say that it’s becoming

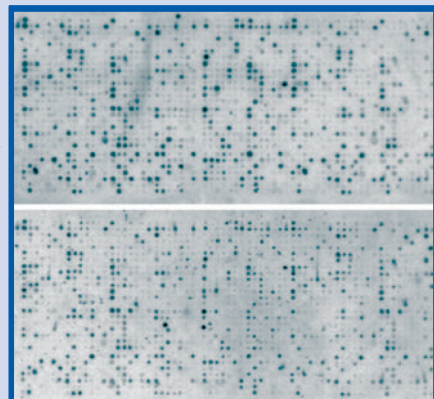
increasingly difficult to be totally comprehensive in signal transduction,” says Konrad Howitz, director of molecular biology at BIOMOL in Plymouth Meeting, Pennsylvania. BIOMOL, now in its twenty-first year of operation, was one of the first companies to specialize in this area and to use signal transduction as a theme for a catalogue.

BIOMOL’s recent product development has been aimed at developing enzymes, substrates, antibodies and inhibitors for the study of histone deacetylases, which play a role in the regulation of gene transcription and other biological processes involving chromatin, and sirtuins, which are protein deacetylases implicated in the regulation of ageing. BIOMOL recently merged with UK-based Affiniti Research Products, and now offers the Affiniti line of ubiquitin and proteasome research tools.

“We still find that the model of niche excellence is the one that appeals to us and our customer base,” says Ian Ratcliffe, president and chief operating officer of Upstate in Charlottesville, Virginia, another established company supplying cell-signalling tools. Ratcliffe says Upstate’s customers are evenly split between academia and industry.

“We think of ourselves as a ‘content’ company,” says Ratcliffe, and so is always on the lookout for companies with novel detection technologies. One such is CIS bio, based in Bagnols-sur-Cèze, France, which sells reagents and assays based on time-resolved homogenous fluorescence resonance energy transfer (HTRF) for probing molecular interactions. The company’s proprietary HTRF technology uses a luminescent europium (Eu^{3+})-based cryptate compound as donor to tag one of the interacting partners and a phycobiliprotein (XL665) as acceptor to tag the other. When the two partners

PEPSCAN: DATA COURTESY OF J. TUYNMAN & D. RICHEL
ACADEMIC MEDICAL CENTER, THE NETHERLANDS



PepChip protein kinase profiles of normal (top) and tumour tissue.

burgh, Pennsylvania, and the Opera high-throughput confocal microscopy platform from Evotec of Hamburg, Germany.

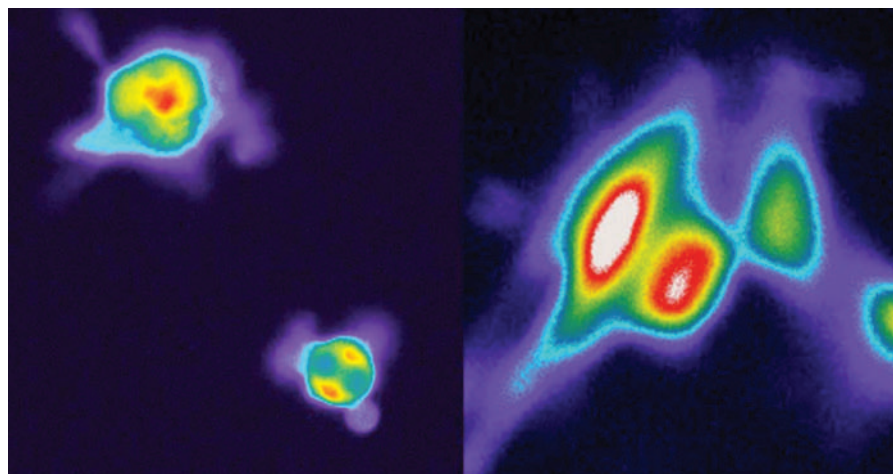
Blocking on cue

The intracellular interaction between the GPCR and its associated G protein is also a potential drug target. This is the approach of drug-discovery company cue BIOTech of Evanston, Illinois, which is also marketing its Minigene vectors. These deliver small peptides that block signal transduction through a specific G protein by competing for the site where the G protein would normally bind to the receptor. "You can turn off each G protein individually inside the cell," says Annette Gilchrist, president and founder of cue, which licensed the technology from Northwestern University. The company offers both plasmid-based cDNA vectors for transient transfection and retroviral vectors for hard-to-transfect cells.

Pure and simple

Many potential drug targets, such as GPCRs and ion channels, are transmembrane proteins, and their isolation and study is difficult as they often need to remain embedded in a lipid membrane to maintain their structural integrity.

By exploiting the fact that cell-membrane proteins are incorporated into the surface envelope of budding retroviruses, Integral Molecular of Philadelphia, Pennsylvania, has developed a way of isolating membrane proteins from cells while preserving their native structure. Cells producing non-infectious retroviral particles are also engineered to make high levels of the desired membrane protein. As the virus core buds from the cell it



Cells treated with a cue BIOTech Gq Minigene vector (left) respond less strongly than normal cells (right) to stimulation by thrombin.

is surrounded by cell membrane enriched with the protein. The resulting Lipoparticles marketed by the company are non-infectious spheres (100–150 nm in diameter) of retroviral core protein surrounded by a phospholipid bilayer containing around 100 molecules of the desired membrane protein in its native conformation. The particles are stable when frozen and refrigerated. "Generally, the preparations that we make are greater than 100 picomoles of membrane protein per milligram of total protein," says Benjamin Doranz, president, chief scientific officer and founding partner of Integral Molecular.

"The most complex protein that we've obtained is a 14-spanning membrane protein amino-acid transporter," says Doranz. But there will be limits, he says, such as membrane proteins with extremely long

cytoplasmic tails that interfere with viral assembly or those that do not traffic to the cell surface.

Lipoparticles can serve as a source of homogeneous and structurally intact membrane proteins for high-throughput screening, monoclonal antibody production and structure analysis using X-ray crystallography. They are also being paired with optical biosensors, such as the surface plasmon resonance (SPR)-based detectors developed by Biacore, of Stockholm, Sweden, for the kinetic analysis of membrane protein interactions with antibodies and ligands.

Characterizing kinases

Protein kinases are garnering increasing attention as drug targets, particularly in the light of the recent success of Gleevec, a tyrosine kinase inhibitor developed by Swiss-

Continued from page 243

bind, fluorescence resonance energy transfer (FRET) can be induced between the cryptate and XL665, which emits a long-lived fluorescence at 665 nm. CIS bio entered into partnerships with both Upstate and antibody company Cell Signaling Technology of Beverly, Massachusetts, in August this year — a move that will extend the reach of its HTRF technology in the life sciences. CIS Bio also recently linked up with detection-system manufacturers BMG LABTECH of Offenburg, Germany, Molecular Devices of Sunnyvale, California, and Tecan of Zurich, Switzerland, to develop HTRF-compatible fluorimeters or upgrade kits for optimizing existing equipment. Upstate is also collaborating with Dharmacon of Lafayette, Colorado, to develop small interfering RNA (siRNA)-based products as an alternative to gene knock-outs, and with Evident Technologies on the quantum dot front (see 'Quantum dots begin to show their true colours', page 247).

In September, Serologicals Corporation, which also owns Chemicon of Temecula, California,



ZeptoSens cell lysate array.

announced plans to acquire Upstate. Ratcliffe sees it as a good fit, combining the company's strengths in phosphorylation, nuclear function, drug discovery and screening with Chemicon's expertise in developing tools for neurobiology. "There's little overlap but very complementary ranges of products," he says.

Companies such as Amersham Biosciences, now part of GE Healthcare, aim to offer integrated solutions. "We're looking at this holistically and we're not interested in just providing a small part of the process. We're interested in providing innovative solutions that go across the biology, the hardware and software and into the informatics," says John Anson, vice-president of development at GE Healthcare Biosciences. These products range from image analysers for cell-based assays (IN Cell Analyzer 1000 and 3000 and LEADseeker) to pH-sensitive cyanine dyes for tracking the internalization of cell-surface receptors and GFP-based live-cell translocation assays developed in collaboration with BioImage of Soeborg, Denmark.

ZEPTOSENS

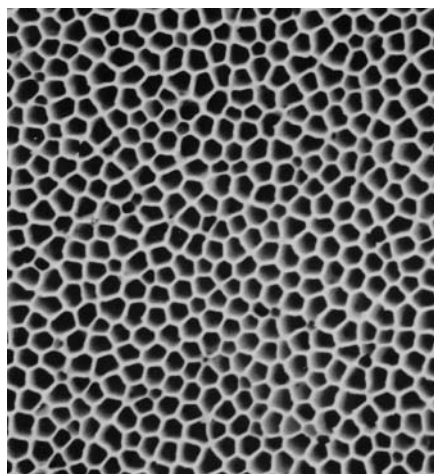
D.G.

based pharmaceutical company Novartis, which was approved for the treatment of chronic myelogenous leukaemia in 2001. Gleevec represents a new class of drugs that disrupt components of the intracellular signalling pathways that cause cells to grow and divide uncontrollably, giving rise to cancers.

There are more than 500 protein kinase genes in the human genome. They catalyse the specific phosphorylation of proteins and play an essential role in many signalling pathways, including those involved in cell-cycle control.

One way of evaluating the substrate selectivity and function of protein kinases in a high-throughput format is by peptide microarrays to which the kinases will bind selectively. Pepscan Systems of Lelystad, the Netherlands, launched its PepChip Kinase peptide microarray about a year ago. "Currently, we have 1,150 different substrates," says Jos Joore, vice-president of array technology at Pepscan, all derived from a public database. The microarray can be used for substrate profiling of known and unknown kinases and for specificity testing of kinase inhibitors. The company is also turning its attention to proteases. Longer term, Joore says, it is trying to improve the specificity of its peptide substrates by providing them as constrained loops, or a combination of constrained loops, rather than linear peptides.

The three-dimensional surface of the PamChip from PamGene of 's-Hertogenbosch, the Netherlands, is designed to allow



The porous PamChip surface provides a greater area for protein immobilization.

peptide substrates to be deposited at higher concentrations than conventional arrays. "We can immobilize much more material per square millimetre than other flat materials," says Rob Ruijtenbeek, PamGene's head of kinase research. A 500-fold increase in reactive surface compared with two-dimensional arrays is claimed.

At the heart of the system is PamGene's 5D-Pulse flow-through microarray technology, in which peptides are covalently immobilized using inkjet technology onto the porous microarray surface via the peptide amino terminus. Sample is then pumped back and forth through the porous material to facilitate

mixing. The pumping-cycle frequency can be changed and detection is with fluorescent antibodies using a CCD camera/microscope. The system provides kinetic readouts in which substrate conversion is monitored over time; traditional array formats limit detection to a single time point.

In June, PamGene announced plans to join forces with Jerini Peptide Technologies (JPT) in Berlin, Germany, which allows PamGene to marry its microarray platform with JPT's comprehensive peptide sets for kinase profiling. Zeptosens of Witterswil, Switzerland, is similarly increasing the range of its arrays by selling antibodies developed by Cell Signaling Technology of Beverly, Massachusetts, for use with the Zeptosens planar waveguide detection protein microarray platform. With this technology, only fluorophores located at or near the surface of the waveguide are excited and signals from unbound molecules in the bulk solution are not detected.

According to Peter Oroszlan, director of business development at Zeptosens, this technology provides a significant increase in signal-to-background ratios, enabling detection of low-abundance proteins such as signalling molecules. "Our system allows you to measure 600 protein molecules in a spot, which corresponds to one zeptomole [10^{-21} moles]," he says. ZeptoMARK protein microarrays are available in capture or reverse-array formats (see *Nature* **429**, 102; 2004), and applications include expression monitoring of proteins during drug profiling and pathway mapping,

PAMGENE

QUANTUM DOTS SHOW THEIR TRUE COLOURS

As an alternative to organic dyes, quantum dots (QDs) may be worth a closer look. These semiconductor nanometre-sized crystals, typically with a cadmium-based core, avoid some of the shortcomings associated with traditional organic dyes and fluorescent proteins (see *Nature* **413**, 450; 2001). QDs are brighter, not prone to photobleaching, and come in a wide range of colours. By changing the size of the particle, the emission can be tuned to any



Evident Technologies' quantum dots come in many colours.

wavelength, from ultraviolet to infrared. QDs also have narrow emission spectra, which means more colours can be used at a time with minimal channel overlap. Multiple colours of QDs can be simultaneously excited using a single light source. QDs have, for example, been used to track the activity and diffusion of individual glycine receptors in the neuronal membrane for up to 20 minutes — considerably longer than is possible with Cy3 dyes.

"We don't expect QDs to completely displace organic dyes," says Clint Ballinger, chief executive of Evident Technologies in Troy, New York. "They make an attractive alternative for many different applications." Evident began by making ultrafast optical switches for the telecom industry. But when the bottom fell out of that business, the company turned to alternative applications for its nanocrystals. It sells its QDs (EviDots) as EviTags, which are water-stabilized, conjugation-ready QDs that can be coupled to proteins or antibodies, and as EviFluors, which have biotin groups.

Another company with know-how in this area is Quantum Dot Corporation (QDC) of Hayward, California. Its product line includes basic Qdot nanocrystals, Qdot bioconjugates coupled to proteins, oligonucleotides and small molecules (streptavidin, protein A and biotin), and Qbead microspheres. In June, QDC introduced a kit that enables researchers to conjugate their own antibodies to Qdot nanocrystals — a procedure that the company says typically takes 3–4 hours. The kit can also be used to couple other thiol-containing molecules to QDs.

So what of the future? "Some customers don't like cadmium," says Ballinger, and Evident plans to launch indium phosphide-based EviDots in the next few months as an alternative to cadmium for the visible range.

D.G.

monitoring activation-state markers, such as phosphorylated proteins, and the study of disease progression.

Read-out

Biosensors for tagging proteins to monitor their activation levels and distribution continue to get more sensitive and versatile. The impact of green fluorescence protein (GFP) and its variants in studying cell signalling is immense. Biosensors that use GFP can, for example, detect conformational changes in proteins in response to ligand binding, changes in protein localization or changes in protein activity.

New fluorescent dyes are also being developed. Klaus Hahn and his colleagues at the University of North Carolina School of Medicine, Chapel Hill, recently reported a biosensor that can visualize the natural dynamics of an unlabelled endogenous intracellular signalling protein, the GTPase Cdc42, in living cells. The sensor is composed of the Cdc42-binding fragment of the Wiskott–Aldrich syndrome protein (WASP), to which a novel merocyanine dye has been coupled. The dye is sensitive to changes in hydrophobicity that occur at the interface between the interacting proteins. Binding of WASP to GTP-activated Cdc42 causes the dye to fluoresce. The sensitivity provided by direct excitation of a novel fluorescent dye enables detection of protein activation at native levels.

One of the more novel detection platforms is quantum dots (see ‘Quantum dots show their true colours’, page 247); another is



Lighting-up protein interactions: Klaus Hahn (left) and Alexei Touthkine with the dye on which their biosensor is based.

single-molecule photon stamping spectroscopy, which is being used to study the dynamics of the interactions of single proteins (see ‘Probing real-time protein interactions’).

LI-COR Biosciences of Lincoln, Nebraska, offers two-colour near-infrared fluorescence detection of signal transduction events. The firm originally applied its infrared detection technology to western blots, “but a western blot is not particularly convenient for looking at a pathway”, says Michael Olive, LI-COR’s vice-president of research and development. The company has developed the In-Cell Western assay for quantifying proteins in fixed cultured cells in 96- or 384-well microplates in less time than conventional western blots by

bypassing the need for lysate preparation and the use of gels and membranes. The use of two spectrally distinct near-infrared dyes effectively doubles the number of endpoints that can be analysed, enabling, for example, the measurement of both phosphorylated extracellular signal-regulated kinases (ERKs) and total ERK protein at the same time.

With slight tweaking, the ‘In-Cell Western’ can become the ‘On-Cell Western’. This was developed by James Wager-Miller in the department of anaesthesiology at the University of Washington in Seattle, who is using it to follow the internalization and recycling of GPCRs, in particular the cannabinoid receptor 1, to and from the cell surface. The hope is that this will lead to a better understanding of how these trafficking events can lead to the desensitization of cells following prolonged or repeated exposure to agonists.

The In-Cell Western is amenable to automation and this month the company will launch a new two-colour plate reader called Aeriis, which automates the assay. This may take the technology “into the realm of lead validation”, says Olive, helping to prevent costly drug failures later on.

The next challenge for cell signalling will be to look at cellular behaviour on a global scale and for this further improvements in technology will be needed, along with better computational and mathematical tools for deriving information about complete signal-transduction networks. ■

Diane Gershon is assistant editor, *Nature Medicine* Technical Reports.

PROBING REAL-TIME PROTEIN INTERACTIONS

Interactions between proteins are one of the main ways in which signals are transferred onward in intracellular signalling pathways. But few methods of studying these interactions can capture the dynamic nature of molecular recognition within signalling complexes.

A group led by Peter Lu, senior research scientist in the W.R. Wiley Environmental Molecular Sciences Laboratory at Pacific Northwest National Laboratory in Richland, Washington, has now developed a method to do just that.

Using a technique called single-molecule photon stamping spectroscopy, which detects and analyses photons that are emitted as single proteins interact, the group has been able to capture proteins in motion as they flip-flop against each other in a manner somewhat analogous to line flicking in fly-fishing. “This type of behaviour cannot be measured by static structure measurements such as X-ray crystallography or NMR,” says Lu.

His team has used this approach to study the interaction of single molecules of activated Cdc42 with the dye-labelled WASP (Wiskott–Aldrich syndrome protein) biosensor developed in collaboration with Klaus Hahn and his colleagues (see main text above). Their results

indicate highly dynamic — rather than static — interactions between this pair of signalling proteins.

Lu’s set-up is largely ‘home-built’. Measurements of the interaction of single molecules of Cdc42 with the dye-tagged WASP fragment are made using confocal fluorescence microscopy with laser excitation.



Peter Lu studies the dynamic interactions of signalling complexes.

Fluorescence photons are directed towards a photon-stamping detector, which records the intensity and duration of the photon peaks; the data are then analysed to determine the dynamics of the protein–protein interaction.

Technical limitations at present are photobleaching of the dye molecule by the excitation laser and movement of the complex away from the laser focal point. Lu is experimenting with methods for confining the complex either using agarose gel or by tethering one of the protein partners to a glass surface.

Lu plans to extend the single-molecule spectroscopy approach to other important biomolecular complexes under physiological conditions, ultimately hoping to study single-molecule protein conformational dynamics in living cells. “Right now, the most interesting part may not be the technique but the new scientific information we learn,” he says.

D.G.

Company	Products/activity	Location	URL
Products and services for signal-transduction research			
Alexis Biochemicals	Suppliers of reagents for cell biology, signal transduction and molecular biology	Lausanne, Switzerland	www.alexis-corp.com
Assay Designs	MAP kinase ELISAs, Antibody Shop antibodies	Ann Arbor, Michigan	www.assaydesigns.com
Avanti Polar Lipids	Phospholipids and sphingolipids	Alabaster, Alabama	www.avantilipids.com
BD Biosciences Pharmingen	Catalogue and custom antibodies; kits for immunoassays	San Diego, California	www.pharmingen.com
Bender Medsystems	ELISA systems for research and <i>in vitro</i> diagnosis; instant ELISAs for cytokines	Vienna, Austria	www.bendermedsystems.com
Biocept	3D HydroArrays for gene expression related to apoptosis and signal transduction	Carlsbad, California	www.biocept.com
Biolmage	Cell-based assays and pathway screening for drug discovery	Soeborg, Denmark	www.bioimage.com
BIOMOL	Biochemicals for life-science research including signal transduction	Plymouth Meeting, Pennsylvania	www.biomol.com
BioSource	Custom peptides; off-the-shelf and customized antibodies for signal-transduction research	Camarillo, California	www.biosource.com
Calbiochem	Products for signal transduction, protein chemistry and glycobiology	San Diego, California	www.calbiochem.com
Cell Signaling Technology	SignalSilence validated siRNA kits; antibodies, including motif-specific antibodies; proteins and kits for cell-signalling research	Beverly, Massachusetts	www.cellsignal.com
Chemicon	Catalogue and custom mono- and polyclonal antibodies; growth factors and cytokines	Temecula, California	www.chemicon.com
CIS bio International	HTRF (homogeneous time-resolved fluorescence)/TRACE reagents and kits for time-resolved-FRET assays for monitoring molecular interactions	Bagnols-sur-Cèze, France	www.htrf-assays.com
cue BIOTech	Minigene vectors for probing receptor–G-protein interactions	Evanston, Illinois	www.cuebiotech.com
Cytomx	Cell lines and cDNAs for kinases, receptors and ion channels	Cambridge, UK	www.cytomx.com
DiscoverX	Cell-based assays for protein kinases and second messengers	Fremont, California	www.discoverx.com
Evident Technologies	EviTags and Biotin EviFluor quantum dot nanocrystalline labels for molecular interaction assays	Troy, New York	www.evidenttech.com
GE Healthcare	Instruments, equipment and products for signal transduction assays and cellular assays	Little Chalfont, UK	www.amersham.co.uk
GloboZymes	Signal transduction enzymes, proteins and antibodies	Carlsbad, California	www.globozymes.com
Integral Molecular	Lipoparticles for isolation of membrane proteins	Philadelphia, Pennsylvania	www.integralmolecular.com
JPT Peptide Technologies	Kinase substrate identification service; PepSPots for high-speed epitope mapping; peptide substrates for phosphatase profiling and protease profiling	Berlin, Germany	www.jerini.com
LI-COR Biosciences	Infrared imaging system for western blot and in-cell western blot assay analysis	Lincoln, Nebraska	www.licor.com
Mabtech	Catalogue monoclonal antibodies for cytokine detection in ELISpot, ELISA and intracellular staining, including anti-monkey monoclonal antibodies to interleukins	Nacka Strand, Sweden	www.mabtech.com
Molecular Devices	Liquid-handling and microplate-processing equipment, imaging instruments and assay systems; IMAP substrate finder for protein kinases	Sunnyvale, California	www.moleculardevices.com
Molecular Probes	Fluorescent probes; fluorescence-tagged proteins, substrates and reagents for signal transduction research	Eugene, Oregon	www.probes.com
Norak Biosciences	Fluorescence-based assay for G-protein-coupled receptors for drug discovery	Research Triangle Park, North Carolina	www.norakbio.com
OriGene Technologies	Human full-length cDNA clones, including complete GPCR, protein kinase, and nuclear hormone receptor full-length cDNA collections	Rockville, Maryland	www.origene.com
Panomics	TranSignal Human Cytokine Antibody Arrays	Redwood City, California	www.panomics.com
Pamgene	5D-Pulse Microarray platform for DNA, protein and peptide microarrays	's-Hertogenbosch, the Netherlands	www.pamgene.com
PatoBios	Fluorescence-based assay for G-protein-coupled receptors for drug discovery	Toronto, Ontario	www.patobios.com
Pepscan Systems	Peptide synthesis; epitope mapping; custom and off-the-shelf PepChip kinase substrate peptide arrays for kinase profiling	Lelystad, the Netherlands	www.pepscan.com
ProKinase	Validation of new protein kinases as drug targets; recombinant protein kinases	Freiburg, Germany	www.prokinase.com
Quantum Dot Corporation	Quantum dot nanocrystalline fluorescent tags for antibody imaging	Hayward, California	www.qdots.com
Sigma-Aldrich	Reagents and biochemicals for life-sciences research; antibodies; Panorama Ab microarray for cell-signalling proteins	St Louis, Missouri	www.sigmaaldrich.com
SuperArray Bioscience	GEArray Signal Transduction Gene Arrays; application-specific cDNA	Frederick, Maryland	www.superarray.com
Stressgen	Reagents and kits for cellular stress, apoptosis, oxidative stress and neurobiology research	Victoria, British Columbia	www.stressgen.com
Tocris Cookson	Chemicals for life-science research, including signal transduction agents; custom synthesis and contract research services	Avonmouth, UK	www.tocris.com
Upstate Group	KinaseProfiler specificity testing, IC50Profiler Express and services for crystallography assay development, bulk protein production and purification; enzymes and antibodies for signal-transduction research; chromatin immunoprecipitation kits	Charlottesville, Virginia	www.upstate.com
Zeptosens	SensiChip DNA microarray systems; ZeptoMARK protein chips with planar optical waveguide technology	Witterswil, Switzerland	www.zeptosens.com
Zymed	Antibodies, ELISA kits, ISH kits and immunohistopathology	South San Francisco, California	www.zymed.com
Imaging			
Alpha Innotech	Alpha Array microarray reader; FluorChem gel documentation systems	San Leandro, California	www.alphainnotech.com
Applied Scientific Instrumentation	Automated systems for microscopy and fluorescence screening	Eugene, Oregon	www.asiimaging.com
Axon Instruments	GenePix 4000B microarray scanner and GenePix Pro microarray-analysis software; ImageExpress high-throughput system for live-cell assays	Foster City, California	www.axon.com
Carl Zeiss	plate::explorer and plate::screen automated systems for ultra-high-throughput screening; microscopes, laser-scanning microscopes, confocal imaging	Göttingen, Germany	www.zeiss.com
Cellomics	ArrayScan HCS and KineticScan HCS systems for automated cell-based high-content screening including molecular translocation to measure multiple signal transduction pathways	Pittsburgh, Pennsylvania	www.cellomics.com

Company	Products/activity	Location	URL
Chroma Technology	Optical filters and fluorescence filter sets	Brattleboro, Vermont	www.chroma.com
Dynex	Absorbance and luminescence microplate readers, microplate washers and automated workstations	Chantilly, Virginia	www.dynextechnologies.com
Fujifilm	Imaging systems for protein electrophoresis and general array analysis	Stamford, Connecticut	www.fujimed.com
GeneFocus	Confocal laser scanning and imaging applications for fluorescence-based assays in genomics and proteomics	Waterloo, Ontario	www.genefocus.com
Kodak	Imaging systems and image-analysis software; Gel Logic digital gel-imaging systems	Rochester, New York	www.kodak.com/US/en/health/scientific/products/family.shtml
Leica Microsystems	DN digital microscope range, microscopes and microscope systems	Wetzlar, Germany	www.leica-microsystems.com
Nikon Instruments	Microscopes; COOLSCOPE digital microscope with screen	Kanagawa, Japan	www.nikon-instruments.jp/eng
Olympus	MIC-D digital microscope, upright and inverted microscopes, stereo microscopes, Fluoview confocal laser scanning microscopes	Melville, New York	www.olympus.com
UVP	Bioimaging systems for visualization, documentation and analysis of gels, membranes and microplates	Upland, California	www.uvp.com
Applied Precision	arrayWoRx biochip reader; cellWoRx for imaging living cells	Issaquah, Washington	www.appliedprecision.com
General			
Abgent	Polyclonal and monoclonal antibodies; customized antibody development; off-the-shelf and custom proteins and peptides; protein expression services	San Diego, California	www.abgent.com
Active Motif	TransAM ELISA kits for transcription factor assays; antibodies; peptide nucleic acids for gene silencing; kits and reagents for cell fractionation; nucleic acid isolation	Carlsbad, California	www.activemotif.com
Affymetrix	DNA microarrays, scanners	Santa Clara, California	www.affymetrix.com
Agilent	2100 Bioanalyzer for genomics and proteomics	Palo Alto, California	www.agilent.com
Analytikjena	Instrumentation for molecular spectroscopy, affinity measurements; FLASHScan plate readers; BIAffinity instrument for measuring binding affinity	Jena, Germany	www.analytikjena.com
Apogent Technologies	Labware and equipment for life sciences	Portsmouth, New Hampshire	www.apogent.com
Applied BioPhysics	Automated instruments for measuring electrical impedance in cells	Troy, New York	www.biophysics.com
Applied Biosystems	Protein sequencers, mass spectrometers, chromatography systems, peptide synthesizers, ICAT stable isotope labelling reagents	Foster City, California	www.appliedbiosystems.com
BD Biosciences	FACS range of flow cytometers; antibody arrays for proteomics; reagents and biochemicals for molecular biology	Franklin Lakes, New Jersey	www.bd.com
Beckman Coulter	Automated tools for molecular biology, biochemistry, genomics and proteomics, ProteomeLab	Fullerton, California	www.beckmancoulter.com
Biacore	Analysis and measurement of biomolecular interactions using surface plasmon resonance	Uppsala, Sweden	www.biacore.com
BIOBASE	TRANSFAC family of databases and analysis tools for gene expression, promoters and signalling pathways; contract bioinformatics services	Wolfenbüttel, Germany	www.biobase.com
Biogenesis	Antibodies and immunoassay-related reagents; bulk polyclonal and monoclonal antibodies; custom antibody development and conjugation service	Poole, UK	www.biogenesis.co.uk
Bio-Rad	Products, instruments and software for life-sciences research; Bio-Plex system for multiplexed antibody-type assays	Hercules, California	www.bio-rad.com
Biosite	Diagnostic antibodies; Omniclonal phage-display system for human Ab library development	San Diego, California	www.biosite.com
BMG LABTECH	Microplate and array readers and handling systems	Offenburg, Germany	www.bmg-labtechnologies.com
Cambio	STAR FISH chromosome paints; specialized biochemicals	Cambridge, UK	www.cambio.co.uk
CyBio	Pipetting, liquid handling, incubation and imaging systems for automated screening; Cellight complete system for cell-based screening; CyBio Nanoscan microplate reader	Jena, Germany	www.cybio-ag.com
Dharmacon	siRNA kits and arrays; custom RNAs; large-scale RNA synthesis	Lafayette, Colorado	www.dharmacon.com
Evotec OAI	Services for drug discovery; compound libraries, target assays, medicinal chemistry	Hamburg, Germany	www.evotecoi.com
Hamilton Company	MicroLab automated liquid-handling workstations	Reno, Nevada	hamiltoncomp.com
Harvard Apparatus	Instruments and equipment for electrophysiology and cell biology	Holliston, Massachusetts	www.harvardapparatus.com
Invitrogen	Kits and reagents for cloning, genomics, proteomics, molecular and cell biology; antibodies, fluorescent-tagged antibodies	Carlsbad, California	www.invitrogen.com
Luminex Corporation	xMAP platform for microbead-based multiplex assays	Austin, Texas	www.luminexcorp.com
MTR Scientific	Markets a range of products for life-sciences research from small biotech suppliers	Ijamsville, Maryland	www.mtrscientific.com
Multi Channel Systems	Automated equipment for <i>Xenopus</i> injection, ion-channel screening; systems for studying cultured nerve cells and neuroregeneration	Reutlingen, Germany	www.multichannelsystems.com
PerkinElmer Life Sciences	UltraView confocal laser scanning microscope system for live cell imaging; HydroGel BioChip	Boston, Massachusetts	las.perkinelmer.com
Pierce Chemical/ Pierce Biotechnology	Components and consumables for automated systems in molecular biology, protein chemistry, sample handling, immunology and chromatography	Rockford, Illinois	www.piercenet.com
Promega	Vectors, reagents and kits for genomics, proteomics and cellular analysis	Madison, Wisconsin	www.promega.com
Qiagen	BioRobot multifunctional workstation; reagents, kits and instruments for genomics and proteomics	Venlo, the Netherlands	www.qiagen.com
Roche Diagnostics	Reagents and kits for molecular biology, functional genomics and proteomics research	Lewes, UK	www.roche-applied-science.com
Tecan	Automation for the life sciences; microplate readers for HTRF-based assays	Medford, Massachusetts	www.tecan.com
Thermo Shandon	Antibodies, immunohistochemical reagents, general laboratory equipment	Runcorn, UK	www.thermo.com
Wako Chemicals	Speciality chemicals, bioproducts and clinical diagnostic reagents	Richmond, Virginia	www.wakousa.com

● see advertisement

Contacts

Publisher: Ben Crowe
Editor: Paul Smaglik
Marketing Manager: David Bowen

US Head Office, New York

345 Park Avenue South, 10th Floor,
New York, NY 10010-1707
Tel +1 800 989 7718
Fax +1 800 989 7103
e-mail: naturejobs@natureny.com

US Sales Manager/ Corporations:

Peter Bless

Classified Sales Representatives:

Phone: +1 800 989 7718

Pennsylvania, New York,

Maryland/ Latin America/ NIH:

Michael McGillion

Midwest USA:

Wade Tucker

East USA/ Canada:

Janine Taormina

San Francisco Office

Classified Sales Representative:

Michaela Bjorkman

West USA/ West Corp. Canada

225 Bush Street, Suite 1453

San Francisco, CA 94104

Tel +1 415 781 3803

Fax +1 415 781 3805

e-mail: m.bjorkman@nature.sf.com

European Head Office, London

The Macmillan Building, 4 Crinan Street,
London N1 9XW, UK

Tel +44 (0) 20 7843 4961

Fax +44 (0) 20 7843 4996

e-mail: naturejobs@nature.com

Naturejobs Sales Director:

Nevin Bayoumi (4978)

Production Manager: Billie Franklin

To send materials use London
address above.

Tel +44 (0) 20 7843 4814

Fax +44 (0) 20 7843 4996

e-mail: naturejobs@nature.com

Naturejobs web development:

Tom Hancock

Naturejobs online production:

Stefan Hales

European Satellite Office

Germany/ Austria/ Italy/

The Netherlands/ Belgium:

Patrick Phelan, Odo Wulffien

e-mail: p.phelan@nature.com

o.wulffien@nature.com

Japan Head Office, Tokyo

MG Ichigaya Building (5F),

19-1 Harakatamachi,

Shinjuku-ku,

Tokyo 162-0841

Tel +81 3 3267 8751

Fax +81 3 3267 8746

Asia-Pacific Sales Director:

Rinoko Asami

e-mail: rasami@naturejpn.com

naturejobs

The secrets of success

Grant applications often fail — it's a fact of life. But the deliberations that lead to rejection are not made public, much to the relief of the applicants. The value of breaking this vow of secrecy, however, was made clear at the end of last month at a meeting held in Toronto, Canada, by the American Society of Human Genetics.

To give junior scientists an insight into the grant-reviewing process, the meeting featured a mock assessment of applications. Decisions are made by weighing up five factors: significance, approach, innovation, investigator and environment. And the panel, which featured actual grant reviewers grading fictional proposals, aimed to show how these factors come into play.

First up was a proposal to analyse genetic variation based on small deletions. All of the panellists agreed that the approach was technologically innovative, scientifically sound and would produce quality data. But Vishwajit Nimgaonkar, a geneticist from the University of Pittsburgh, Pennsylvania, scored the proposal low on the basis of significance. "It's not clear to me why we need another set of polymorphisms," he said. "It looks like an application in search of a use." Another panellist had some questions about the 'environment' — in this case whether a key collaborator was really committed to the project.

At the end of the session, the panellists debriefed the audience. As in the real world, they said, many of the applicants were able to cover all five factors adequately in their proposals — but they hadn't, assuming instead that the reviewers would be able to read their minds. In fact, the panellists had repeatedly been lenient to new scientists, saying that not spelling out how the proposal met the five criteria was a typical 'young investigator error'. One senses that the young scientists present all hoped that real reviews are undertaken in a similar atmosphere of understanding.

Paul Smaglik

Naturejobs editor



Contents

SPECIAL REPORT

Getting on the fast-track
to full-time employment p254

WWW.NATUREJOBS.COM

Career centre
Information on the
scientific job market

FOCUS

SPOTLIGHT

RECRUITMENT

ANNOUNCEMENTS

EVENTS

The road to your first full-time position can be long and tortuous. But some researchers have found a shortcut to success. Eugene Russo reports.

David Liu had his future planned. It was May 1999, and the bioorganic chemist had just received his doctorate from the University of California, Berkeley. Liu was all set to return to his undergraduate home at Harvard University to pursue a generously funded, independent postdoc when his career took an unexpected leap.

Out of the blue, Liu began to get interview invitations for tenure-track positions at top chemistry departments including those at Berkeley, the California Institute of Technology and even Harvard. "It still remains a little bit of a mystery," he says. Liu quickly abandoned the idea of a postdoc and instead became a principal investigator at Harvard. Today,

Liu says he is pleased that he declined a postdoc to give his career an early boost. He has enjoyed the responsibility — and is even content with the sometimes onerous duties of managing a lab, and writing grants and papers.

But Liu's experience is unusual. Long training stints and relatively few permanent academic positions can mean lengthy delays for scientists looking to secure their first permanent job. Supply readily exceeds demand for many disciplines, leaving a lot of researchers languishing in postdocs as they scour the circuit

for their first position as principal investigator (M. Teitelbaum *The Public Interest* No. 153, 40–53; 2003).

A recent survey by the American Chemical Society (ACS), illustrates the present trends within the marketplace. Around half of the respondents said that they planned to do a postdoc — in good economic times, that figure can be as low as 30%, says Jura Viesulas, manager of employment information at the ACS. Annual surveys by the American Institute of Physics paint a similar picture: 68% of the respondents from the class of 2003 were planning to do a postdoc, up from 43% during the boom years of the early 1990s. And in the life sciences things are not much better. Data collected by the National Research Council and the US National Institutes of Health suggest that life scientists tend not to get their first grants until their mid-to-late 30s, rather than in their late 20s as they did in the 1970s — a hold-up primarily caused by extended training periods as graduate students and postdocs (see *Nature*

422, 354–355; 2003).

Delays in getting a first full-time job cost scientists money. In 2001, the American Society for Cell Biology released a report suggesting that, compared with engineering, medicine or business, over the course of their career a bioscientist can lose more than a million dollars in opportunity costs, the loss of salary and retirement investments because of long stints as graduate students and postdocs. Clearly young scientists — especially young biomedical scientists — need to plan ahead and keep multiple career options open or they will run the risk of losing time, money and job security.

BREAKING FREE

So how can they break free of the trap? There are several ways to minimize — or even skip completely — the time spent as a postdoc. These include leaving the academic world, emphasizing teaching at liberal arts institutions rather than working for large research institutions, and being geographically flexible.

Astronomer Andrea Schweitzer took the route away from academia. As an undergraduate, she had an idealized vision of the astronomy professor teaching attentive students while sitting under a tree on campus. When that image was shattered with the 'publish or perish' reality, she chose to forgo postdoc and teaching opportunities in favour of permanent employment as an engineer and assistant project manager at high-tech firm Honeywell in Fort Collins, Colorado.

Schweitzer, who is now an independent consultant and who also chairs the American Astronomical Society's employment committee, offers some simple tips to those seeking to expedite their science training and career. Think at least two jobs



Head start: David Liu became a principal investigator without doing a postdoc.

FAST TRACK:
Principal investigators

"An assistant professorship is such a cherished part of being a scientist, that when faced with one, it's really hard to turn it down." — David Liu



D. RIDLEY/CORBIS

ahead, she says. Share an office with a colleague who is one to two years ahead of you, and choose a research topic that is unlikely to be delayed by external events, such as the launch of a spacecraft. And when it comes to putting yourself in a position to apply for a job, “done is better than perfect”, Schweitzer adds.

LEARNING CURVE

In contrast, Karl Haushalter opted to stay in academia — but was a postdoc for just two years before he found his ideal job. The path that led Haushalter to his present position as an assistant professor of chemistry and biology at Harvey Mudd College in Claremont, California, began when family considerations encouraged him to move from Cambridge, Massachusetts, to southern California. After spending a year as a guest lecturer at the University of California, Irvine, Haushalter began a postdoc at the University of California, San Diego. By that time, he knew that he wanted to teach — preferably at a liberal arts college. Halfway through his postdoc project, he saw an opening at Harvey Mudd and jumped at the opportunity. “I felt like this was too good to pass up,” he says.

For those who are interested in moving their career more towards teaching, Haushalter recommends gaining teaching experience as a graduate student (see *Naturejobs* 5; 28 February 2002). He also advises young scientists to ensure they have the skills they need before coming to a small liberal arts college. Unlike working at a large research university, advice from colleagues doing similar research may not be available ‘just down the hall’.

Flexibility is important for those who hope to move quickly from training to a full-time position, says Sally McArthur, a lecturer in biomedical engineering at the University of Sheffield, UK. After receiving her doctorate in Australia, McArthur travelled halfway around the world to the University of Washington, Seattle, for a postdoc. Two years later, she began looking for jobs. After some initial disappointment, she found her current position with the help of a supportive mentor and some informative networking at international meetings. She runs a new biomedical engineering programme for undergraduates, a position with an unusual amount of responsibility and authority for a faculty rookie. She is still adjusting — McArthur’s administrative duties have supplanted any sustained focus on her own research, something she hopes will soon change.

IN SEARCH OF FUNDS

Nic Tapon, a lab head at the charity Cancer Research UK in London, also crossed the globe to follow opportunities that matched his personal and professional priorities. He began his biomedical education in Britain with a four-year graduate degree at University College London. After heading to Boston for a four-year postdoc, he found a job as a research scientist at the University of Nice in his native France. But although his position was secure, funding was scarce and bureaucracy commonplace. “Like a lot of French scientists, I found the lack of funding for science quite frustrating,” says Tapon. “That was definitely a reason to move on.” (see *Nature* 428, 108 and 430, 283; 2004). He returned to Britain for a well-funded position at Cancer Research UK.

For those looking for a programme that could accelerate their education and lead to permanent employment, Michael Teitelbaum, a demographer with the Alfred P. Sloan Foundation in New York, recommends trying one of 100 ‘professional’ master’s degree programmes that the Sloan Foundation has helped start since 1997. Located at universities across the United States, the degrees include training not only in science, but also in law, management and communication. Early indications, Teitelbaum says, are promising. Successful graduates’ salaries are reasonable, and job placement good. But students usually have to pay their own way through the course, a major drawback. It will be several more years before the programme can be judged an unqualified success, Teitelbaum notes.

Pushing to accelerate your career does have a downside, says Liu. When one of his graduating students came for advice on whether he should accept an offer to go directly to a full-time faculty position at Brigham Young University in Provo, Utah, Liu hesitated. Getting additional skill sets and having time to establish an interesting and original research programme during a postdoc could be the key to success, he said. And there is no rush to learn how to manage a lab or to write grants and papers. But despite Liu’s advice, the student took the job. “Maybe that’s a reflection that it’s a very hard thing to turn away,” says Liu. “An assistant professorship is such a fundamentally exciting and cherished part of being a scientist, that when faced with one, it’s really hard to turn that down.” ■

Eugene Russo is a freelance science writer in Takoma Park, Maryland.